A THESIS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Molecular Biological Characterization of Mosquitocidal Strain, *Bacillus thuringiensis* subspecies *mogi*

모기 살충성 strain, *Bacillus thuringiensis* subspecies *mogi* 의 분자 생물학적 특성 연구

By

Qin Liu

Department of Agricultural Biotechnology

Seoul National University

February, 2014

Molecular Biological Characterization of Mosquitocidal Strain, *Bacillus thuringiensis* subspecies *mogi*

Major in Entomology

Department of Agricultural Biotechnology, Seoul National University

Qin Liu

ABSTRACT

Bacillus thuringiensis subspecies *mogi* was isolated from fallen leaves, sampled in a forest region of the city of Mungyeong, Korea. Plasmids from this species have been implicated in pathogenicity as they carry genes responsible for a variety of entomo-pathogenic diseases. The purpose of this study was to characterize the *B. thuringiensis* subsp. *mogi* strian, determine the full genome sequence, and investigate the molecular genetics of expression of novel toxin-related *cry* genes which located on the plasmid in *B. thuringiensis* subsp. *mogi*.

As a primary study, the flagellated vegetative cells of *B. thuringiensis* subsp. *mogi* were agglutinated with the H3 reference antiserum and further agglutinated with 3b and 3d monospecific antisera but non-reactive to 3c and 3e factor sera. These results

create a new serogroup with flagellar antigenic structure of 3a3b3d, designated serovar *mogi*. *B. thuringiensis* subsp. *mogi* showed activity against dipteran larvae, *Culex pipiens molestus* and *Culex pipiens pallens* while no lepidopteran toxicity. It produced three small ovoidal-shaped parasporal crystals combined together and whose SDS-PAGE protein profile consisted of several bands ranging from 75 to 30 kDa. Through the identification of the protein by nano-LC-ESI-IT MS analysis, the putative peptides of Cry27Aa, Cry39ORF2, and Cry20-like were detected. In contrast to the complicated plasmid profiles of *B. thuringiensis* H3 serotype strains, the *B. thuringiensis* subsp. *mogi* contained only megaplasmids (> 30 MDa) on which the toxin genes were occasionally located.

Second, full genome sequence of the novel *B. thuringiensis* subsp. *mogi* strain was determined. The 6.0 Mb genome of *B. thuringiensis* subsp. *mogi* contains three replicons: a circular chromosome (5.40 Mb) encoding 5,652 predicted open reading frames (ORFs) and two mega-plasmids, pMOGI364 (364,564 bp) and pMOGI222 (222,348 bp). The G+C contents of these replicons ranged from 31.3% to 34.2% for pMOGI364 and pMOGI222, respectively, and did not significantly deviated from that of the chromosome (35.2%). There were seventeen toxin-related genes existed in these two mega-plasmids, and six of them (*cry19Bb1, cry73Aa* with *cry40orf2, cry20Bb1, cry27Ab1, cry4Aa* and *cry56Ba1* with *cry39orf2*) belonged to the group of three-domain *cry* toxins.

Finally, to investigate the role of six novel three-domain *cry* genes in crystal production of *B. thuringiensis* subsp. *mogi*, the transcription level of these toxin genes were analyzed by quantitative PCR (qPCR). The results clearly indicated that all of these *cry* genes were successfully transcribed in wild type *B. thuringiensis* subsp. *mogi* strain in different transcription time with different maximum levels. Then, these *cry* genes were cloned to the *Escherichia coli-B. thuringiensis* shuttle vector, pHT1K, under the control of its own promoter, and introduced into an acrystalliferous *B. thuringiensis* Cry$^-$B strain for further molecular characterization. Another vector p1KSD, which containing a strong chimeric *cyt1Aa* promoter combined with the STAB-SD sequence was constructed and used to over-express the *cry* genes. To determine the function of the *cry39orf2* and over-express the *cry56Ba1* in *cry56Ba1* operon, different combinations of Cry56Ba1 and Cry39ORF2 were synthesized in strain Cry$^-$B. The stable inclusion in recombinant cells suggests that Cry39ORF2 assists in synthesis and crystallization of Cry56Ba1 by functioning like the C-terminal domain characteristic of Cry protein in the 130 kDa mass range. In addition, the increased Cry56Ba1 yield under the *cyt1A-p*/STAB-SD promoter has broadened the possibility of application in other toxins.

Key words: *B.thuringiensis,* ovoidal-shaped crystals, mosquitocidal, full genome sequence, three-domain *cry* gene, over-expression

Student number: 2007-23602

TABLE OF CONTENTS

ABSTRACT----------i

TABLE OF CONTENTS ----------iv

LIST OF TABLES ----------vi

LIST OF FIGURES ----------viii

LIST OF ABBREVIATIONS----------xi

LITERATURE REVIEW----------xii

Chapter 1. Characterization of a novel serogroup *Bacillus thuringiensis* strain, subsp. *mogi,* flagellar serotype 3a3b3d

ABSTRACT ----------1

INTRODUCTION----------2

MATERIALS AND METHODS----------3

RESULTS ----------15

DISCUSSION----------25

Chapter 2. Genome sequencing strategy and sequence analysis of *Bacillus thuringiensis* subsp. *mogi*

ABSTRACT ----------28

INTRODUCTION----------29

MATERIALS AND METHODS----------32

RESULTS ----------34

DISCUSSION----------72

Chapter 3. Molecular cloning and characterization of mosquitocidal protein genes from *B. thuringiensis* subsp. *mogi*

ABSTRACT ----------74

INTRODUCTION----------75

MATERIALS AND METHODS----------77

RESULTS ----------95

DISCUSSION----------141

LITERATURE CITED----------186

ABSTRACT IN KOREAN----------195

ACKNOWLEDGEMENTS----------198

LIST OF TABLES

Table 1. Composition of culture media for a new strain of *B. thuringiensis*----------5

Table 2. H agglutination test results of *B. thuringiensis mogi* strain----------7

Table 2-1. Monospecific antisera agglutination test results of *B. thuringiensis mogi* strain----------8

Table 3. Nucleotide sequences of primers used for amplification of the specific probe in southern blot----------14

Table 4. Toxicity of *B. thuringiensis* against *Culex pipiens molestus* and *Culex pipens pallens* 4th instar larvae----------24

Table 5. Read status of *B. thuringiensis* subsp. *mogi* genome assemblies----------36

Table 6. General characteristics of chromosome and gene prediction----------37

Table 7. General characteristics of 2 plasmids from strain *B. thuringiensis* subsp. *mogi*----------37

Table 8. DNA topoisomerase III in *B. thuringiensis mogi.*----------54

Table 9. Nucleotide sequences of forward and reverse primers used for amplification of the target *cry* genes in reverse transcription-polymerase chain reaction (RT-PCR)----------80

Table 10. Nucleotide sequences of primers used for amplification of the target *cry* genes in qPCR----------82

Table 11. Nucleotide sequences of primers used for amplification of the target *cry* genes to pHT1K vector in PCR----------85

Table 12. Primers used for amplification of *cry56Ba1* and *39orf2*----------91

Table 13. Summary of toxin-related genes in pMOGI364 and pMOGI222----------98

Table 14. Amino acid sequences of putative crystal proteins of *B. thuringiensis* subsp. *mogi* strain----------99

Table 15. Toxicity of recombinant strains of *B. thuringiensis* Cry-B against *Culex pipiens molestus* and *Cluex pipiens pallens* 4th instar larvae----------133

Table 16. Toxicity of recombinant strains of *B. thuringiensis* Cry-B against *Culex pipiens molestus* and *Cluex pipiens pallens* 4th instar larvae----------139

Table S1. The CDS in pMOGI364 and their annotations----------146

Table S2. The CDS in pMOGI222 and their annotations----------170

LIST OF FIGURES

Fig. 1. Plasmid patterns from *B. thuringiensis* H3 type strains----------16

Fig. 1-1. Plasmid patterns from *B. thuringiensis* mosquitocidal type strains----------17

Fig. 2. Plasmid pattern from *B. thuringiensis* subsp. *mogi* strain
was examined in pulse field gel electrophoresis----------18

Fig. 3. Transmission electron microscopy of the parasporal crystal
of *B. thuringiensis* subsp. *mogi* strain----------21

Fig. 4. SDS-PAGE and nano-LC-ESI-IT MS analysis----------22

Fig. 5. Southern hybridization of *B. thuringiensis* subsp. *mogi*----------23

Fig. 6. GC-skew analysis of chromosome----------38

Fig. 7. Mummer analysis of *B. thuringiensis mogi* strain with the
selected *Bacillus* species----------39

Fig. 8. Phylogenetic relationship of the *B. thuringiensis mogi*
among the *Bacillus* clade----------42

Fig. 9. Circular representation of pMOGI364----------49

Fig. 10. GC-skew of the plasmid pMOGI364----------51

Fig. 11. Mummer analysis of pMOGI364 with pG9842_209----------52

Fig. 12. Graphic view of pMOGI364 megablast results----------53

Fig. 13. Previews about topoisomerase III in *B. thuringiensis mogi*----------55

Fig. 14. Sequence alignment analysis of the truncated DNA topoisomerase III
with the intact one from NCBI database----------56

Fig. 15. Circular representation of pMOGI222----------64

Fig. 16. GC-skew of the plasmid pMOGI222----------66

Fig. 17. Graphic view of pMOGI222 megablast results----------67

Fig. 18. Alignment of the predict Rep protein of pMOGI222, Rep165 protein
of plasmid pBMB165, Rep63A protein of plasmid pAW63, RepS protein
of pXO2, and RepE protein of pAMβ1 of these three plasmids----------68

Fig. 19. Alignment of the origin of replication in pMOGI222, pBMB165, pBT9727, pAW63, pXO2, pAMβ1 and p43----------70
Fig. 20. The pMOGI222 replication region----------71
Fig. 21. Comparison and clustering of *B. cereus* group plasmids based on replication----------72
Fig. 22. Cloning and expression analysis of mosquitocidal *cry* genes from *B. thuringiensis* subsp. *mogi*----------86
Fig. 23. Schematic illustration of the different constructs containing *cry56Ba1* or *cry39orf2* ----------90
Fig. 24. Upstream region and schematic illustration of constructs used to synthesize Cry proteins----------91
Fig. 25. The over-expression vector p1KSD----------93
Fig. 26. Nucleotide sequence and deduced amino acid sequence of the *cry19Bb1* gene----------100
Fig. 27 Nucleotide sequence and deduced amino acid sequence of the *cry73Aa* and *cry40orf2* gene----------103
Fig. 28. Nucleotide sequence and deduced amino acid sequence of the *cry20Bb1* gene----------107
Fig. 29. Nucleotide sequence and deduced amino acid sequence of the *cry27Ab1* gene----------110
Fig. 30. Nucleotide sequence and deduced amino acid sequence of the *cry4Aa* gene----------113
Fig. 31. Nucleotide sequence and deduced amino acid sequence of the *cry56Ba1* and *cry39orf2* gene----------116
Fig. 32. Comparison of the deduced amino acid sequence of Cry19Bb1, Cry20Bb1, Cry73Aa with other Cry proteins in five conserved blocks----------121
Fig. 33. Comparison of the deduced amino acid sequence of Cry4Aa, Cry27Ab1,

Cry56Ba1 with other Cry proteins in five conserved blocks----------123
Fig. 34. Agarose gel electrophoresis analysis of *cry* genes mRNA obtained by RT-PCR from *B. thuringiensis* subsp. *mogi* in the sporulation stage----------125
Fig. 35. Transcription-level analysis of the six *cry* genes from *B. thuringiensis* subsp. *mogi* were analyzed----------126
Fig. 36. Agarose gel electrophoresis analysis of PCR products----------129
Fig. 37. SDS-PAGE analysis of the recombinant *B. thuringiensis cry*⁻*B* strains containing different *cry* genes----------130
Fig. 38. Phase contrast micrographs of the recombinant *B. thuringiensis* Cry-B strains containing different *cry* genes----------131
Fig. 39. Transmission electron microscopy of the acrystalliferous *B. thuringiensis* Cry-B strain and the recombinant strains----------132
Fig. 40. Over-expression of *cry56Ba1* operon and functional analysis of *cry39orf2*----------137
Fig. 41. Alignment of the ORF2 amino acid sequence with C terminal regions of selected Cry proteins----------140

LIST OF ABBREVIATIONS

BHI: brain heart infusion

bp: base pair(s)

ca.: approximately

DIG: dioxigenin

DNA: deoxyribonucleic acid

dNTP: dexoxyribonucleoside triphosphate

EDTA: ethylenediaminetetra acetic acid

g: acceleration due to gravity

GYS: glucose-yeast extract salt medium

h: hour(s)

kb: kilo base pair(s)

kDa: kilo Dalton

LB: Luria-Bertani medium

min: minute(s)

M: molarity (= mol/l)

OD: optical density

PAGE: polyacrylamide gel electrophoresis

PVDF: polyvinylidene fluoride

LITERATURE REVIEW

Basic biology of *Bacillus thuringiensis*

B. thuringiensis Berliner was originally discovered in Japan over a century ago by Shigetane Ishiwata (Federici *et al*., 2010) as the cause of the sudden ("sotto") death disease of silkworms, larvae of the silkworm moth, *Bombyx mori*. Ten years later, the German bacteriologist Ernst Berliner, unaware of Ishiwata's paper, described a similar bacterium as the cause of disease in larvae of the flour moth, *Ephestia kuhniella*. The species name "*thuringiensis*" is derived from Thuringia, the German state where the diseased flour moth larvae were found.

The gram-positive bacterium *B. thuringiensis*, which can be readily isolated from a variety of environmental sources including soil, water, plant surfaces, grain dust, dead insects, and insect feces (Federici, 1999), was characterized by its ability to produce crystalline inclusions during sporulation. Its life cycle is simple. When nutrients and environmental conditions are sufficient, the spore germinates producing a vegetative cell that grows and reproduces by binary fission. Cells continue to multiply until one or more nutrients, such as sugars, amino acids, or oxygen, become insufficient for continued vegetative growth. Under these conditions, the bacterium sporulates producing a spore and parasporal body, the latter, composed primarily of one or more insecticidal proteins in the form of crystalline inclusions (Federici *et al*., 2010). These

are commonly referred to in the literature as insecticidal crystal proteins or δ-endotoxins, which can compose as much as 40% of the dry weight of a sporulated culture. These inclusions consist of proteins exhibiting a highly specific insecticidal activity (Aronson *et al*., 1986). Most crystal proteins are active against larvae of certain members of the Lepidoptera, but some show toxicity against dipteran (flies) or coleopteran (beetles) insects, or nematodes.

Insecticidal proteins in *B. thuringiensis*

There are two types of insecticidal crystal proteins in *B. thuringiensis*, Cry (for crystal) and Cyt (for cytolytic) proteins, and variations on each of these types. Individual Cry toxins have a defined spectrum of insecticidal activity, usually restricted to a few species within one particular order of insects. To date, toxins for insect species in the orders Lepidoptera (butterflies and moths), Diptera (flies and mosquitoes), Coleoptera (beetles and weevils) and Hymenoptera (wasps and bees) have been identified. The Cry proteins are classified on the basis of amino acid sequence homology, where each protoxin acquired a name consisting of the mnemonic Cry (or Cyt) and four hierarchical ranks consisting of numbers, capital letters, lower case letters and numbers (e.g. Cry25Aa1), depending on its place in a phylogenetic tree. The know Cry and Cyt proteins now fall into 32 sets including Cyt1, Cyt2 and Cry1 to Cry 67 (Crickmore *et al.*, 2010).

The structural diversity of Cry toxin

In *B. thuringiensis*, these proteins form crystals, with the most common types being composed of Cry1 proteins of about 135 kDa. These are primarily toxic to lepidopterous insects, and consist of a N-terminal half (contains the active protein) containing the toxic portion of the molecule, released after ingestion by insect midgut proteases, and a C-terminal half important to crystallization (Schnepf *et al*., 1998). In addition to 135 kDa proteins, Cry proteins of 65-70 kDa are known which correspond to the N-terminal half of the 135 kDa Cry type. Examples include Cry2A toxic to lepidopterous and dipterous insects, Cry3A (similar in mass to Cry2) toxic to coleopterous insects, and Cry11A toxic to certain dipterous insects. Phylogenetic studies indicate that all of the above Cry types evolved over millions of years from the same ancestral molecule, the diversity in host spectra being selected for when mutant strains wound up in the midguts of insect species belonging to different orders.

Five highly conserved blocks exist in the toxic core of most known Cry protoxins, which are important for their activities and specificities (Höfte and Whiteley, 1989). They are arranged in three distinct domains (I–III, from N- to C-termini). Block 1, encompassing the central helix α5 of domain I, has been implicated in pore formation, a role that might explain its highly conserved nature (Gazit *et al.*, 1998). Block 2 includes the C-terminal half of helix α6 and all of α7 of domain I, and the first β-strand of domain II. Helix α7 serves as a binding sensor to initiate the structural

rearrangement of the pore-forming domain (Gazit and Shai, 1995). Residues within block 2 are involved in formation of salt bridges, which could be considerable, in conformational changes upon binding of the toxin to receptor or for maintaining the protein in globular form (Schnepf *et al.,* 1998). Block 3 contains the last β-strand of domain II and the N-terminal segment of domain III, the latter forming the interface with domains I and II. Block 4 corresponds to the second β-strand of domain III that affects the structural integrity of the protein, oligomeric aggregation, and the appropriate function of the ion channels. The highly conserved block 5 in domain III is at the C-terminus of the activated toxin and is another major element that stabilizes the mature toxin (Nishimoto *et al.,* 1994; Yamagiwa *et al.,* 1999)

Endogenous proteases in *B. thuringiensis*

During the early sporulation phase, an increase in intracellular protease activity occurs in *B. thuringiensis* cultures. Proteases endogenous to *B. thuringiensis* have been described from the cysteine, metallo, and serine families of enzymes. Major proteases in most *B. thuringiensis* species are thermostable and many are metalloproteases, with some exceptions. The endogenous proteolytic activities in *B. thuringiensis* may hydrolyze crystal proteins. For example, a reduction in the size of *B. thuringiensis* subsp. *tenebrionis* inclusion crystal proteins (ICPs) occurred during sporulation, and proteolysis was prevented by the addition of protease inhibitors (Carroll *et al.*, 1989). ICPs from *B. thuringiensis* subsp. *kurstaki* crystals, incubated in

denaturing and reducing conditions, were hydrolyzed by metalloproteases in the crystal (Kumar and Venkateswerlu, 1997). Interestingly, the toxin produced under these conditions was highly active against the cotton leafworm, *Spodoptera littoralis*, a species insensitive to native *kurstaki* crystals or toxins generated by exogenous proteases (Kumar and Venkateswerlu, 1998a). Although the crystal contained multiple Cry proteins, the toxin was homogenous, as demonstrated by two-dimensional polyacrylamide gel electrophoresis, and lacked the first 29 amino acids of the protoxin N-terminus (Kumar and Venkateswerlu, 1998b). Mosquitocidal ICPs were also degraded in the crystal (Dai and Gill, 1993). The mosquitocidal protoxin Cry11Aa1 was partially processed from 72 to 32 – 40 kDa proteins within the crystal by endogenous *B. thuringiensis* proteases (Ibarra and Federici, 1986).

Chapter 1. Characterization of a novel serogroup *Bacillus thuringiensis* strain, subsp. *mogi*, flagellar serotype 3a3b3d

ABSTRCT

Bacillus thuringiensis strain *mogi* was isolated from fallen leaves, sampled in a forest region of the city of Mungyeong, Korea. The flagellated vegetative cells of *B. thuringiensis* strain were agglutinated with the H3 reference antiserum and further agglutinated with 3b and 3d monospecific antisera but non-reactive to 3c and 3e factor sera. These results create a new serogroup with flagellar antigenic structure of 3a3b3d, designated serovar *mogi*. The strain *mogi* showed activity against dipteran larvae, *Culex pipiens molestus* and *Culex pipiens pallens* while no lepidopteran toxicity. It produced three small ovoidal parasporal crystals combined together and whose SDS-PAGE protein profile consisted of several bands ranging from 75 to 30 kDa. Through the identification of the protein by nano-LC-ESI-IT MS analysis, the putative peptides of Cry27Aa, Cry39ORF2, and Cry20-like were detected. In contrast to the complicated plasmid profiles of *B. thuringiensis* H3 serotype strains, the *B. thuringiensis* subsp. *mogi* contained only megaplasmids (> 30 MDa) on which the toxin genes were occasionally located. The new type strain, *B. thuringiensis* subsp. *mogi* (H3a3b3d) will be a good resource for novel mosquitocidal *cry* genes.

Key words: *Bacillus thuringiensis*, novel serogroup, mosquitocidal, *cry* genes

1. INTRODUCTION

The Gram-positive and endospore-forming bacterium *Bacillus thuringiensis*, which is frequently used in industrial applications, is well known for its ability to produce crystalline parasporal inclusions that have insecticidal activity against various species. The parasporal inclusion, which may contain more than one type of insecticidal crystal protein (ICPs), is released with the spore upon lysis of the sporangium (Höfte and Whiteley, 1989; Schnepf *et al*., 1998). A number of isolates of the bacterium are commercially produced, with activity against Lepidoptera, Diptera and Coleoptera.

B. thuringiensis produces parasporal inclusions (crystals) having several unique features including insecticidal, nematocidal or anti-cancer activity (Ohba *et al.*, 2009; Roh *et al*., 2007). Numerous *B. thuringiensis* isolates have been collected worldwide and some of them have been characterized by various techniques such as biochemical test, H-serotyping, plasmid patterns and *cry* gene contents by PCR analysis (Lecadet *et al.*, 1999; Porcar and Juarez-Perez, 2003; Reyes-Ramirez and Ibarra, 2008). Among them, the classification of *B. thuringiensis* isolates by H-serotyping has been believed as an efficient way since it is based on the stable and specific characters of the flagellar (H) antigen. The H-serotyping, however, has limitations, proving unreliable as a predictor of insecticidal activity. For example, *B. thuringiensis* serovar *morrisoni* (H8a8b) is a collection of heterogeneous pathovars specifically active against Lepidoptera, Coleoptera, or Diptera, and even those with no insecticidal activities

(Park *et al*., 1998). Nevertheless, it is still of great value in discriminating between *B. thuringiensis* strains (Lecadet *et al*., 1999). By the end of 1998, up to 69 different serotypes and 13 sub-antigenic groups, giving 82 serovars, have been involved in H-serotype classification scheme.

B. thuringiensis is widely distributed and recovered from 70% of soil samples from all continents, with Asian samples being an especially rich source; *B. thuringiensis* subsp. *israelensis* and subsp. *kurstaki* are the most common types (Martin and Travers, 1989). It is always desirable to search for a better insecticide against noxious insects (Sezen *et al*., 2010). In this study, a new subserotype *B. thuringiensis* strain, occurring in the H3 serogroup, which has larvicidal activity against *Culex* mosquitoes, was isolated and characterized.

2. MATERIALS AND METHODS

2.1 Bacterial strains and growth media

The *B. thuringiensis* strain *mogi* was isolated from fallen leaves, sampled in a forest region of the city of Mungyeong, Korea, according to the method of Ohba and Aizawa (Ohba and Aizawa, 1978). Other *B. thuringiensis* type strains include in this work were kindly provides by the International Entomopathogenic *Bacillus* Center (IEBC) at the Pasteur Institute, Paris, France. *B. thuringiensis* were grown at 28ºC

with vigorous shaking in SPY medium for plasmid preparation and GYS medium for expression of crystal proteins (Kronstad *et al*., 1983; Li *et al.*, 2002; Nickerson and Bulla, 1974). The LB medium was used as a primary culture of *B. thuringiensis* and in *E.coli* culture for plasmid preparation. Media compositions were described in Table 1. Brain heart infusion (BHI) medium was used to culture competent *B. thuringiensis* cells.

Table 1. Composition of culture media for a new strain of *B. thuringiensis.*

Medium*	Component	% (g/L)
LB	Trypton	1
	Yeast extract	0.5
	Nacl	1
GYS	Glucose	0.1
	Yeast extract	0.2
	K_2HPO_4	0.05
	$(NH_4)_2SO_4$	0.2
	$MgSO_4$	0.002
	$MnSO_4$	0.005
	$CaCl_2$	0.008
SPY	$(NH_4)_2SO_4$	0.2
	K_2HPO_4	1.4
	KH_2PO_4	0.6
	$Na_3C_6H_5O_7 \cdot 2H_2O$	0.1
	$MnSO_4 \cdot 7H_2O$	0.02
	Glucose	0.1
	Yeast extract	0.1

*LB: Luria-Bertani; GYS: glucose-yeast extract salt medium; SPY: Spizizen medium.

2.2 Preparation of H antisera and H agglutination studies

For H-serotype identification of the strain *mogi*, a slide agglutination test was used as described previously (Ohba and Aizawa, 1978). A motility inhibition test (Ishii and Ohba, 1993) was also involved in H-serotyping to confirm the specificity of the reaction. Reference antisera used were: (1) 55 H-antisera against the type strains of *B. thuringiensis* H1-H55 (Lecadet *et al*., 1999) (Table 2), and (2) four monospecific antisera against H-antigenic subfactors 3b, 3c, 3d, and 3e (Ohba and Aizawa, 1989) (Table 2-1). H antisera-antigen agglutination studies were performed using 96 well plates (Roh *et al*., 1996). One hundred microliter of flagellated bacteria suspension, grown at 30ºC to an OD_{600} of 0.7, was mixed in each well with 100 μl of H antiserum which had been diluted 50-fold with saline. Agglutinin was assayed after incubation at 37ºC for 1 h.

Table 2. H agglutination test results of *B. thuringiensis* subsp. *mogi* strain.

H-serotype	Serovar	*mogi*	H-serotype	Serovar	*mogi*
1	*thuringiensis*	-	25	*coreanensis*	-
2	*finitimus*	-	26	*silo*	-
3a3b3c	*kurstaki*	+	27	*mexicanensis*	-
3a3c	*alesti*	+	28a28b	*monterrey*	-
3a3d	*sumiyoshiensis*	+	28a28c	*jegathsan*	-
3a3d3e	*fukuokaensis*	+	29	*amagiensis*	-
4a4b	*sotto*	-	30	*medellin*	-
4a4c	*kenyae*	-	31	*toguchini*	-
5a5b	*galleriae*	-	32	*cameroun*	-
6	*entomocidus*	-	33	*leesis*	-
7	*aizawai*	-	34	*konkukian*	-
8a8b	*morrisoni*	-	35	*seoulensis*	-
8a8c	*ostriniae*	-	36	*malaysiensis*	-
8b8d	*nigeriensis*	-	37	*andaluciensis*	-
9	*tolworthi*	-	38	*oswaldocruzi*	-
10	*darmstadiensis*	-	39	*brasiliensis*	-
11a11b	*toumanoffi*	-	40	*huazhongensis*	-
11a11c	*kyushuensis*	-	41	*sooncheon*	-
12	*thompsoni*	-	42	*jinghongiensis*	-
13	*pakistani*	-	43	*guiyanggiensis*	-
14	*israelensis*	-	44	*higo*	-
15	*dakota*	-	45	*roskildiensis*	-
16	*indiana*	-	46	*chanpaisis*	-
17	*tohokuensis*	-	47	*wratislaviensis*	-
18	*kumamotoensis*	-	48	*balearica*	-
19	*tochigiensis*	-	49	*muju*	-
20a20b	*yunnanensis*	-	50	*navarrensis*	-
20a20c	*pondicheriensis*	-	51	*xiaguangiensis*	-
21	*colmeri*	-	52	*kim*	-
22	*shandongiensis*	-	53	*asturiensis*	-
23	*japonensis*	-	54	*poloniensis*	-
24	*neoleonensis*	-	55	*palmanyolensis*	-

-, no response; +, agglutination.

Table 2-1. Monospecific antisera agglutination test results of *B. thuringiensis* subsp. *mogi* strain.

monospecific antisera*	*mogi*	monospecific antisera	*mogi*
3b	+	3d	+
3c	-	3e	-

-, no response; +, agglutination.

* The result was co-worked with Ohba and Aizawa at Japan.

2.3 Plasmid preparation and PCR

Plasmid DNA was extracted using the alkaline lysis method (Reyes-Ramirez and Ibarra, 2008) including a step involving lysozyme treatment. Each strain was cultured in 50 ml SPY medium to an optical density at 600 nm of 0.8 to 1.0 at 30°C and 250 rpm shaking. Vegetative cells were pelleted at 7000 rpm for 10 min at 4°C. Each pellet was resuspended in 20 ml cold TES buffer (30 mM Tris base, 5 mM EDTA, 50 mM NaCl; pH 8.0 adjusted with 3 N HCl) and centrifuged under the same conditions. Cells were resuspended in 2 ml lysis buffer (TES buffer containing 20% sucrose, 2 mg/ml lysozyme, and 1µl/ml of RNase from a 10 mg/ml stock solution) and incubated at 37°C for 90 min or until more than 90% spheroplast formation was achieved and monitored under a microscope. The spheroplast suspension was supplemented with 3 ml of 8% sodium dodecyl sulfate in TES buffer and incubated at 68°C for 10 min. Then 1.5 ml of 3 M sodium acetate (pH 4.8) was added, and the suspension was incubated at -20°C for 30 min. The suspension was centrifuged at 10, 000 rpm for 20 min at 4°C. The supernatant was translucent; if it was not, another centrifugation was done, and ultimately, if still required, it was filtered. Two volumes of cold absolute ethanol were added to the supernatant and incubated overnight at -20°C. Plasmid enriched DNA was pelleted at 10,000 rpm for 20 min at 4°C. Each pellet was dissolved in 100 µl Tris-EDTA (pH 8.0) (10 mM Tris-HCl, 1 mM EDTA) and stored

at -20°C until further use.

For *cry*-gene typing in the strain *mogi*, PCR tests were done according to the method of Lee (Lee *et al*., 2001). Twenty major *cry*/*cyt* genes primers (*cry1Aa, cry1Ab, cry1Ac, cry1B, cry1C, cry1D, cry1E, cry1F, cry1I, cry2A, cry3A, cry3B, cry3C, cry4A, cry4B, cry7A, cry9A, cry10A, cry11A, and cyt1A*) were synthesized for PCR analysis. The purified PCR products were ligated to pGEM-T easy vector (Promega Co., USA) for sequencing and analyzed by dye termination method in ABI 377 automated sequencer (Applied Biosystems, USA).

2.4 Pulsed-field gel electrophoresis

Separation and examination of large DNA fragments was performed with pulsed-field gel electrophoresis as described by Dean and Bazylinski (1999) in a CHEF-DRII system (Bio-Rad Ltd, Richmond, CA). Agarose gels were prepared at a concentration of 1% and electrophoresis was performed in 0.5 × TBE buffer (45 mmol/l Tris–HCl, 45 mmol/1 boric acid, 2.5 mmol/l EDTA, pH 8.2) at 14°C for 16 h. Mid-Range II PFG Markers (New England Biolabs Ltd, Ipswich, MA) were used as molecular weight markers. The field strength and pulse conditions were 6 V/ cm, and switch times ramped from1 to 25 s.

2.5 Transmission electron microscopy

Crystal morphology of the isolate was examined by phase-contrast microscopy and

transmission electron microscopy. For TEM sample, the *B. thuringiensis* cells were harvested prior to autolysis and washed with sterile water once. Cells were primary fixed with 2 ml fixation buffer (2% paraformaldehyde and 2% glutaraldehyde in 0.05 M sodium cacodylate buffer, pH 7.2) for 2 h at 4ºC, followed by three washes with 0.05 M sodium cacodylate buffer. After post fixation in 1% osmium tetroxide for 2 hours and two brief washes with ddH_2O, the samples were stained with 0.5% uranyl acetate overnight and dehydrated in increasing concentrations of ethanol. Then the specimens were embedded in Spurr's resin at 70ºC for 24 h. Sections were cut with an ultramicrotome (MT-X, RMC, Tucson, AZ, USA) and were stained with 2% uranyl acetate and Reynold's lead citrate. The cells were observed under a transmission electron microscope (Libra 120, Carl Zeiss, Germany).

2.6 SDS-PAGE and MS analysis

Parasporal inclusions were purified by the method of Thomas and Ellar (1983). For SDS-PAGE samples, cell were cultured on NA (nutrient agar) medium plate at 28ºC and harvested after autolysis. SDS-PAGE was performed on a 12% separating gel with 5% stacking gel. The gel was stained with 0.1% Coomassie brilliant blue (Sigma Co., St Louis, MO, USA). The stained protein bands were identified by nano-LC-ESI-IT MS analysis performed by Korea Basic Science Institute. For the

similarity search, we performed BLAST searches with the query of matched sequences found in the contig sequences database of *Bacillus* using the SEQUEST program (version 3.3.1, Thermo Electron Corporation, USA).

2.7 Southern hybridization

Southern hybridization to total plasmid DNA of *B. thuringiensis mogi* strain was performed according to the manufacturer's instruction (Boehringer Mannheim, Germany). Total plasmid DNAs of *mogi* were separated on 0.8% agarose gels. The gels were treated for 15 min in 0.25 N HCl and transferred to Hybond N^+ filters (Amersham Pharmacia, Biotech, Sweden) in 0.2 N NaOH as transfer buffer. PCR amplified *cry* genes (Table 3) from *B. thuringiensis* servoar *mogi* strain were used as probes and labeled with digoxigenin using a DIG DNA labelling kit (Boehringer Manheim Co., Germany). Prehybridization, hybridization, washing and detection procedures were followed as described by the manufactures.

2.8 Insects and toxicity assays

The mosquito larvicidal activities were assayed on 4th instar larvae of *Culex pipiens molestus* and *Culex pipiens pallens* (Diptera: Culicidae) which were grown in a container (35 × 25 × 3 cm) at 25ºC. Freeze-dried *B. thuringiensis* spores-crystal

complex were suspended in double-distilled water. Suspensions were diluted to 6 or 7 different concentrations in cups in a final volume of 100 ml. Bioassays were replicated three times using 30 4th *Culex pipiens* instar of per concentration. After 48 h of exposure at 25°C, dead larvae were counted. Statistical analysis of data and 50% lethal concentrations (LC_{50}) were performed with probit analysis (Russell *et al.*, 1977).

Table 3. Nucleotide sequences of primers used for amplification of the specific probe in southern blot.

Primer [a]	**Sequence (5'-3')**	**Target gene**
Fw-27p	ATGAATCCTTATCAGGATAAGAATGAA	*cry27Aa*
Re-27p	ATTCTGATCGTACGTATTATATCCTT	
Fw-39p	CCGGCTGCACATGTAACC	*cry39orf2*
Re-39p	GGTTACATGTGCAGCCGG	
Fw-20p	ACATGTAGAACAACTTATTCAAC	*cry20-like*
Re-20p	GTTCTAATCCTGAATCCCCTG	

3. Results

3.1 Characteristic of *B. thuringiensis mogi*

According to H-serotyping test, flagellated vegetative cells of the isolate were agglutinated with the H3 reference antiserum only (Table 2). In a further test to identify the subfactors, *mogi* was agglutinated with 3b and 3d antisera but non-reactive for 3c and 3e antisera (Table 2-1). It is clear from the results that the H-antigen of the strain *mogi* comprises three subfactors: 3a, 3b, and 3d. The subfactor 3a is an antigen commonly contained in all of the *B. thuringiensis* strains that belong to the serotype H3 (De Barjac *et al.*, 1981; Ohba and Aizawa, 1989). According to the current H-serotyping scheme for *B. thuringiensis* (Lecadet *et al*., 1999), the serotype H3 is divided into four subserotypes: 3a3c (serovar *alesti*), 3a3b3c (serovar *kurstaki*), 3a3d (serovar *sumiyoshiensis*), and 3a3d3e (serovar *fukuokaensis*). Thus, our present results create a new subserotype, 3a3b3d, designated serovar *mogi*. The serovar name is derived from ‘‘mosquito” in Korean.

The plasmid DNA pattern of the isolate was compared with the profiles of *B. thuringiensis* H3 type strains (Fig. 1) as well as *B*. *thuringiensis* mosquitocidal type strains (Fig. 1-1). In contrast to the complicated plasmid profiles of H3 serotype and mosquitocidal type strains, *B. thuringiensis* subsp. *mogi*, which contains a very simple pattern without visible small plasmid observed in these figures. For further study, pulse field gel electrophoresis was carried out to confirm the result. Fig 2 clearly indicated that, there were at least two megaplasmid (larger more than 194 kb) bands but none small plasmid harbored in this strain.

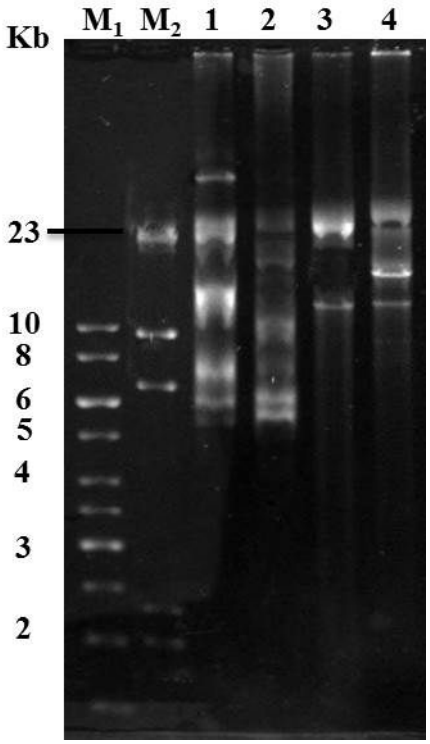


Fig. 1. Plasmid patterns from *B*. *thuringiensis* H3 type strains. The serovars are given, Lanes: 1, serovar *kurstaki* strain HD-1 (serotype 3a,3b,3c); 2, *alesti* (3a,3c); 3, serovar *sumiyoshiensis* (serotype 3a,3d); 4, *fukuokaensis* (3a,3d,3e). M_1, Gene Ruler^{TM} 1 kb DNA ladder; M_2, lambda DNA digested with *Hind* III.

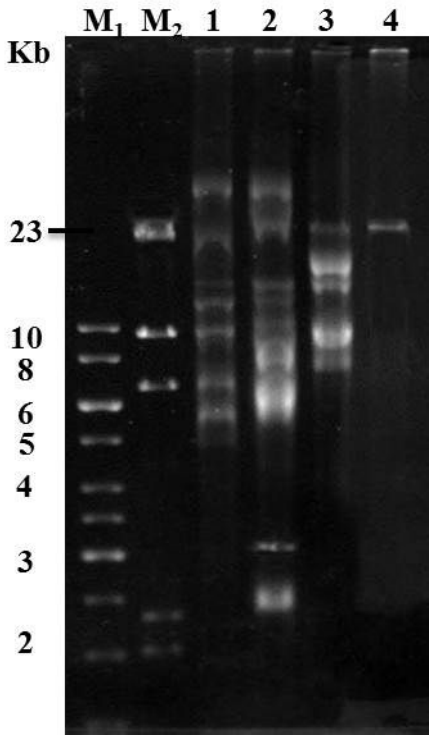


Fig. 1-1. Plasmid patterns from *B. thuringiensis* mosquitocidal type strains. The serovars are given, Lanes: 1, *israelensis* (H14); 2, *kyushuensis* (H11a11c); 3, *morrisoni* PG14 (H8a8b); 4, *mogi* (H3a3b3d). M_1, Gene Ruler^{TM} 1 kb DNA ladder; M_2, lambda DNA digested with *Hind* III.

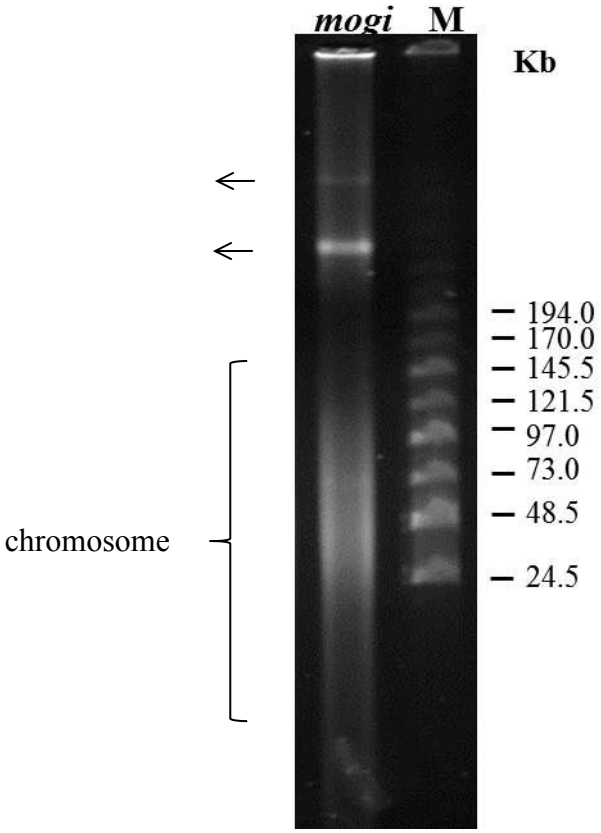


Fig. 2. Plasmid pattern from *B. thuringiensis* subsp. *mogi* strain was examined in pulse field gel electrophoresis. The Mid-Range II PFG Marker (M) was used as molecular weight markers (electrophoresis condition: 1% agarose gel, 6 V/cm, 15 for 18 h, switch times ramped from 1-25 s).

For detection of crystal genes of *B. thuringiensis* strains, PCR analysis was performed with *cry* gene-specific primers. The PCR-based identification of *B. thuringiensis cry* genes was first developed by Carozzi *et al*. (1991), who introduced this technique as a tool for prediction insecticidal activity. The PCR test with 20 specific primers failed to detect the genes allied to *cry1, cry2, cry3, cry4, cry7, cry9, cry10, cry11* and *cyt1*.

As shown in Fig. 3, the parasporal body is composed of 3 major endotoxins, a large ovoidal inclusion and two smaller inclusions combine together, with an distinct membrane outside the inclusions in one sporangium. The SDS–PAGE profile of the crystal proteins consisted of several bands ranging from 20 to 75 kDa (Fig. 4A). Fig 4B shows the internal amino acid sequences of putative Cry proteins obtained by nano-LC-ESI-IT MS analysis. Through the protein identification, three putative peptides of Cry39ORF2, Cry27Aa and Cry20-like were detected. Interestingly, the peptide No. 3 was evident in several fragments (Nos. 2, 3, 4, and 7) and the peptides Nos. 4, 5 and 7 were also found in two fragments. The existence of these putative peptides supports the observation that the mosquito-specific activity was associated with the strain *mogi*. Previous investigators reported the occurrence of mosquitocidal proteins Cry27Aa (Saitoh *et al*., 2000) in a serovar *higo* (H43) strain, and Cry20-like (Lee and Gill, 1997) in a serovar *fukuokaensis* (H3a3d3e) strain. Southern blot analysis showed that all of these three *cry* genes located in the total plasmid DNA of

strain *mogi* (Fig 5), this strain harbored three different *cry* genes at least.

3.2 Toxicity of *B. thuringiensis* subsp. *mogi*

The toxicity of wild-type *B. thuringiensis* subsp. *mogi* was evaluated against 4th instars of *Culex pipiens molestus* and *Culex pipeins pallens* larvae (Table 4). The spore-crystal mixture of the strain gave high mortalities of the two mosquito species, *C. pipiens molestus* (with an estimated LC_{50} of 16 μg/ml) and *C. pipens pallens* (with an estimated LC_{50} of 22.2 μg/ml). In contrast, it exhibited no larvicidal activity against three lepidopteran species: *Bombyx mori*, *Plutella xylostella* and *Lymantria dispar* (data not shown).

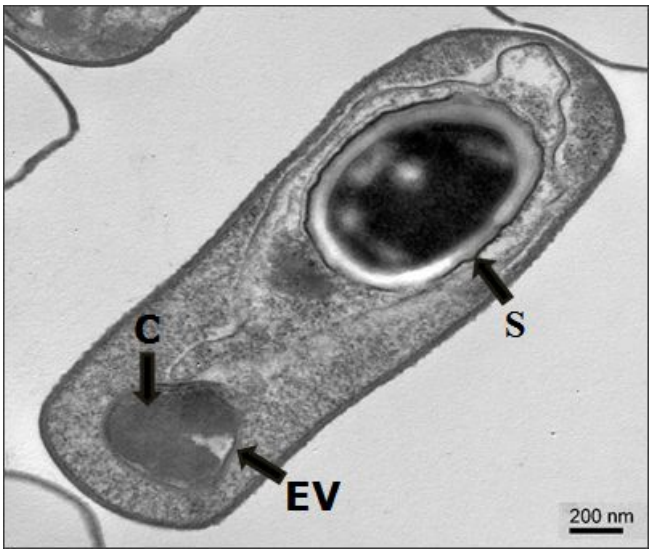
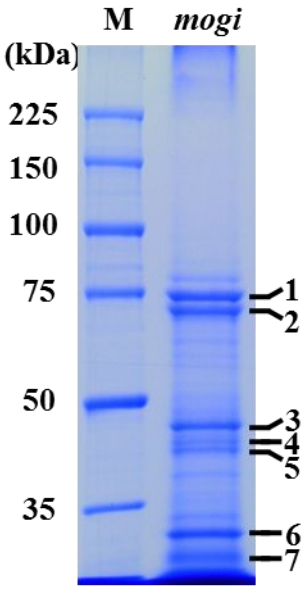


Fig. 3. Transmission electron microscopy of the parasporal crystal of *B. thuringiensis* subsp. *mogi* strain. Panel: C, S and EV indicate parasporal crystal, spore and envelope, respectively. Magnification is 60,000×.

A



B

Peptide No.	Internal sequence of tryptic peptide	Protein fragment No.	Best matched Protein	Predicted molecular mass	Serovar	GenBank Accession no.	Reference
1	K.YPLANDPQMY LR.N	1	Cry27Aa	94 kDa	*higo*	Q9S597	Saitoh et al. (2000)
2	K.TVEVFPESDRV R.I	2	Cry39ORF2	63 kDa	*aizawai*	BAB72017	-
3	R.IMQAYNLYDAR .N	2, 3, 4, 7					
4	K.AQLDGSGLAR .T	2, 3	Cry27Aa	94 kDa	*higo*	Q9S597	Saitoh et al. (2000)
5	R.YVPQISQVPAV K.A	2, 3					
6	K.ITTINLGDYDK. I	2					
7	R.SAATGAIYGIS R.S	5, 7	Cry20 like	86 kDa	*fukuokae nsis*	O32321	Lee and Gill (1997)

Fig. 4. SDS–PAGE and MS analysis of *B. thuringiensis* subsp. *mogi.*

A. SDS–PAGE profile of the parasporal crystal of *B. thuringiensis* subsp. *mogi* strain. Panel: M indicates the molecular marker.

B. Internal amino acid sequences of putative crystal proteins of the *B. thuringiensis* subsp. *mogi* identified by nano-LC-ESI-IT MS analysis.

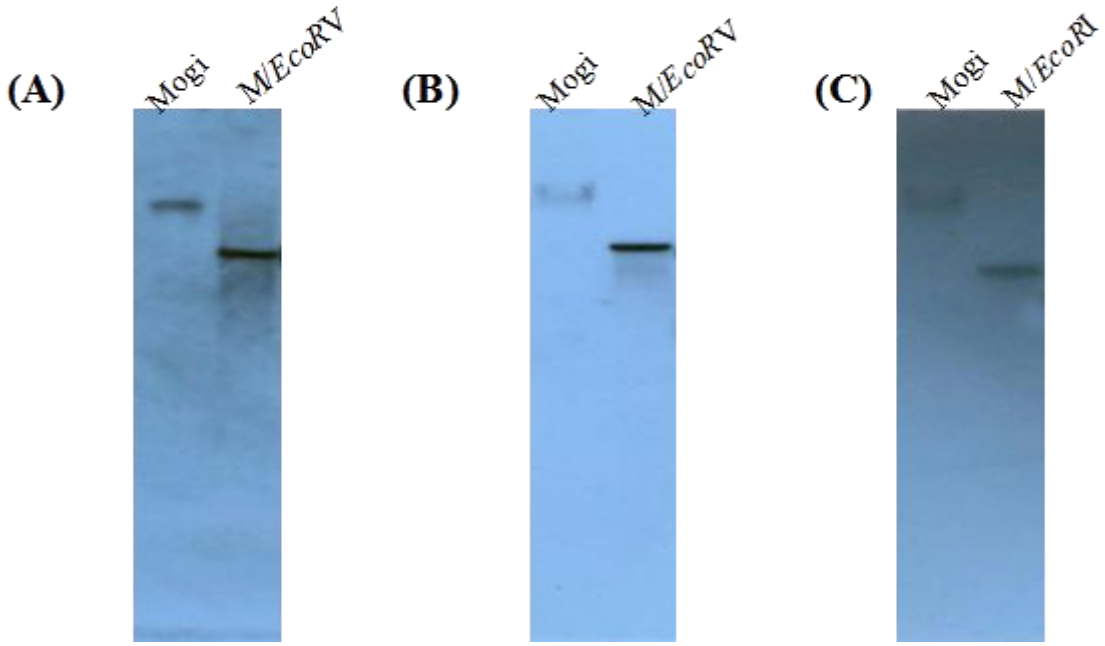


Fig. 5. Southern hybridization of *B. thuringiensis* subsp. *mogi* plasmid DNA with specific regions of *cry27Aa* (A), *cry39orf2* (B), and *cry20-like* (C) as probes , respectively. Lanes: Mogi, *B. thuringiensis* subsp. *mogi* plasmid DNA; M/*EcoR* V, *B. thuringiensis* subsp. *mogi* plasmid DNA digested with *EcoR* V; M/*EcoR* I, *B. thuringiensis* subsp. *mogi* plasmid DNA digested with *EcoR* I.

Table 4. Toxicity of *B. thuringiensis* against *Culex pipiens molestus* and *Culex pipiens pallens* 4th instar larvae.

Strain	***Culex pipiens molestus***		***Culex pipiens pallens***	
	LC_{50}[a] (µg/ml)	FL_{95}[b] (µg/ml)	LC_{50}[a] (µg/ml)	FL_{95}[b] (µg/ml)
mogi	16.0	14.1-21	22.2	13.7-27.5
israelensis	3.4	2.1-5.3	4.0	3.4-6.4

[a]LC_{50}: 50% lethal concentration (in µg) of freeze-dried spore–crystal complex per milliliter after 48 hours. The data are the total of three assays as determined by Probit analysis.

[b]FL_{95}: fiducial limits at P=0.95.

4. Discussion

B. thuringiensis is a gram-positive soil bacterium characterized by its ability to produce parasporal inclusions during sporulation. A number of *B. thuringiensis* isolates of the bacterium are commercially produced, with activity against Lepidoptera, Diptera and Coleoptera. Novel isolates with insecticidal activity have been recovered from numerous sources, particularly soil (Yamamoto and Powell, 1993), as well as grain dusts, diseased insect larvae, animal feed mills, phyloplane and aquatic environments (Coole, 1995). Meanwhile, *B. thuringiensis* shows great variability, as has been demonstrated by the huge number of strains isolated around the world (Xu *et al.*, 2013), by the number of serotypes known to date (a total of 84) (Roh *et al.*, 2009), and by the great number of different *cry* gene sequence accumulated so far (a total of 492), as well as by the number of molecular characterization tools that have been developed, such as sequencing of the *flagellin* gene and of the *gyrB* and *aroE* genes, the band patterns from repetitive extragenic palindromic-PCR analyses, and the plasmid patterns, among others (Choi *et al*., 2012; Heo *et al*., 2012; Koo *et al*., 2012; Reyes-Ramirez and Ibarra, 2008), all indicating the great variability within this species.

In this study, the characterization of a novel serogroup *B. thuringiensis* strain was reported. Characterization was based on serotype, plasmid pattern, crystal inclusion,

Cry protein composition, *cry* gene content and insect toxicity.

The diversity in flagellar H3 antigen agglutination reactions is one indication of the enormous genetic diversity among *B. thuringiensis* isolates. The plasmid pattern from *B. thuringiensis* subsp. *mogi*, showed a much simpler profile than other type strains. Strains of *B. thuringiensis* usually exhibit complex plasmid profiles, with molecules ranging from 2 to more than 200 kb (Hoflack *et al.*, 1997). Plasmids play a crucial role in bacterial evolution and adaptation by mediation the horizontal exchange of genetic material and providing advantageous functions to their carrier. The *B. thuringiensis* subsp. *israelensis,* which is highly toxic to larvae of several dipteran aquatic insects, has been reported to contain up to 10 plasmids (González and Carlton, 1984). Interest has been predominantly focused on large molecule plasmids where most of the crystal protein genes are encoded on. However, some small plasmids have been ascribed as no functions other than maintenance. They are referred to as cryptic plasmid. There are only megaplasmids but no small plasmid harbored in *mogi* strain made the isolate very different.

Sporulation in *B. thuringiensis* is associated with high protease activity which coincides with the onset of crystal formation (Andrews *et al*., 1985). Proteases endogenous to *B. thuringiensis* have been described from the cysteine, metallo, and serine families of enzyme, which can degrade Cry proteins and affect insect toxicity. Carroll *et al* (1989) reported that there was a reduction in the size of *B. thuringiensis*

subsp. *tenebrionis* inclusion crystal proteins occurred during sporulation, and proteolysis was prevented by the addition of proteases inhibitors. Mosquitocidal proteins were also degraded in the crystal (Dai and Gill, 1993).

In general, Cry proteins are active against Lepidopteran (Cry1 of 130–140 kDa), both Lepidopteran and Diptera (Cry II of 71 kDa), Coleopteran (Cry III of 66–77 kDa) and Diptera (Cry IV of 125–145 and 68 kDa) larvae (Guz *et al*., 2005; Salehi *et al*., 2008).

In this study, characteristically, isolate possessed δ-endotoxins with molecular weights between 20 and 75 kDa, among which ~70 kDa was distinctly present signifying their spectrum of activity against Diptera. A few crystalline inclusions were composed of the small components of polypeptides of 20 - 45 kDa. Both the MS analysis and southern blots proved the strain *mogi* contained three Cry proteins (Cry27Aa, Cry39ORF2 and Cry20-like), and the size of these protoxin is around 63 kDa to 94 kDa. It is well known that *B. thuringiensis* produces endogenous proteases and their production may vary considerably among strains (Rukmini *et al*., 2000). The present of these small polypeptides indicated the protoxins maybe cleaved during sporulation phase, giving rise to the smaller size. The new isolate also showed toxicity against *C. pipiens molestus* and *C. pipiens pallens* in bioassay test. In conclusions, the novel serovar type strain, subsp. *mogi* (H3a3b3d), will be a good resource for screening mosquitocidal Cry proteins.

Chapter 2. Genome Sequencing Strategy and Sequence Analysis of *Bacillus thuringiensis* subsp. *mogi*

ABSTRCT

Bacillus thuringiensis belongs to the *Bacillus cereus* sensu lato group as well as *B. anthracis* and *B. cereus*. Plasmids from this group of organisms have been implicated in pathogenicity as they carry genes responsible for a variety of mammalian and entomo-pathogenic diseases. In this study, genome sequence of the novel serogroup of *B. thuringiensis* subsp. *mogi* (H3a3b3d) was determined. The 6.0 Mb genome of *B. thuringiensis mogi* contains three replicons as follows: a circular chromosome (5.40 Mb) encoding 5,652 predicted open reading frames (ORFs) and two mega-plasmids, pMOGI364 (364,564 bp) and pMOGI222 (222,348 bp). The G+C contents of these replicons ranged from 31.3% to 34.2% for pMOGI364 and pMOGI222, respectively, and did not significantly deviated from that of the chromosome (35.2%). About 200 kb sequence of pMOGI364, showed a high similarity (more than 90% identity) to the plasmid pG9842_209 of *B. cereu*s G9842, and the last 146 kb fragment of pMOGI364 was found to harbor nine *cry* genes. The analysis of the replication-related sequence suggests that pMOGI222 may belong to the pAMβ1 family of Gram-positive theta-replicating plasmids. These sequences possibly contribute to the expansion of the pathogenic *B. thuringiensis* plasmid gene pool.

Key words: *Bacillus thuringiensis*, genome sequencing, megaplamid, *cry* gene, theta replicating mode

1. INTRODUCTION

Members of the *Bacillus cereus* group of organisms include *B. cereus, B. anthracis* and *B. thuringiensis*. This group of Gram-positive spore-formers forms a highly homogeneous subdivision of the genus *Bacillus*. The presence of Cry protein crystals in the spore is speculated to give *B. thuringiensis* an advantage in the soil environment upon sporulation (Jensen *et al*., 2003) over *B. cereus*, *B. thuringiensis* is phenotypically distinguished from *B. cereus* only by the formation of intracellular protein crystals during sporulation. Overall, genetic studies have shown that *B. cereus* and *B. thuringiensis* are essentially identical (Helgason *et al*., 1998).

The presence of a complex arrangement of plasmid DNA is a common characteristic of many strains in *B. thuringiensis*. The number and size of these plasmids (2–250kb) vary considerably among strains (González and Carlton, 1980). Often, plasmids confer an obvious advantage to the host, or encode traits that favor their own maintenance and survival. For many plasmids, however, no functions other than maintenance have been ascribed. They are referred to as cryptic plasmids.

In a plasmid pattern, two different groups of plasmids can be recognized: those that

are smaller than 30 MDa and those that are larger 30 MDa, called megaplasmids (Reyes-Ramirez and Ibarra, 2008). Small plasmids are generally present in high copy numbers, while megaplasmids are present in low copy numbers. As for the megaplasmids, their main recognized function is harboring *cry* genes, although the sequencing of some of these plasmids indicates the occurrence of other important genes (Berry *et al*., 2002; Chao *et al*., 2007; Jensen *et al*., 1995).

In addition, small plasmids generally use the rolling-circle replication mechanism, with single-stranded DNA intermediates, while megaplasmids normally use the "theta" replication mechanism (Wilcks *et al*., 1999). Theta replicons are currently divided into six groups (Group A-F) (http://www.essex.ac.uk/bs/staff/osborn/DPR/DPR_ThetaData.htm). Although there have been relatively few studies focusing on the characterization of Gram-positive theta replicons, as opposed to their Gram negative counterparts, plasmids pertaining to the broad host- range pAMβ1 family (group D) have been mostly studied from Gram-positive bacteria (Braund *et al.*, 1993; Brantl *et al.*, 1990; Swinfield *et al.*, 1990).

To date, five plasmids from the *B. cereus* group have been analyzed and reported to belong to the pAMβ1 family. Of these, the largest detected plasmid pBMB165 (about 82 kb) from *B. thuringiensis* subsp. *tenebrionis* YBT-1765, its mini-replicon has been determined (Huang *et al.,* 2006). p43 (65 kb) comes from *B. thuringiensis* subsp. *kurstaki* HD263, and a 2,828 bp replication region of p43 has been cloned (Baum and

Gilbert, 1991). The broad-host-range conjugative plasmid pAW63 (71,777 bp) has been isolated from *B. thuringiensis* subsp. *kurstaki* HD73, and a 4.1 kb replicon of pAW63 has been characterized (Wilcks *et al*., 1999). pBT9727 (77,112 bp) was the sole plasmid in the pathogenic strain *B. thuringiensis* subsp. *konkukian* 97-27, and its replication protein and the predicted origin have been analyzed by sequence comparison (Rasko *et al*., 2005). pXO2 (96,231 bp) was the second virulence plasmid of *B. anthracis*, and a 2,429 bp replication region has been identified (Tinsley *et al.*, 2004).

For *B. thuringiensis* genomics, there are 11 complete and 19 in-progress genomes publicly available on NCBI (http://www.ncbi.nlm.nih.gov/genome/486, as of November 12, 2013). *B. thuringiensis* strains have a genome size of 5.31 to 6.77 Mb. Here, the complete genome from an environmental isolate of *B. thuringiensis* subsp. *mogi* was determined by using shotgun libraries plus paired-end library sequencing strategy. The use of both libraries showed a more adequate representation of contigs and permitted the closure of the genome sequences. Also a comparative analysis with the genome of other *Bacillus* was proceeded and these data provide an insight into evolutionary relationships among the *Bacillus*.

2. MATERIALS AND METHODS

2.1 Bacterial strains and growth media

The novel serogroup *B. thuringiensis mogi* (H3a3b3d) strain used in this research, was isolated from fallen leaves, sampled in a forest region of the city of Mungyeong, as previously described (chapter 1) (Roh *et al.*, 2009). The LB medium was used as a primary culture of *B. thuringiensis* and the second culture of *B. thuringiensis* was grown at 28ºC with vigorous shaking in SPY medium for DNA preparation.

2.2 Plasmid DNA extraction and sequencing strategy

Plasmid DNA of *B. thuringiensis* subsp. *mogi* was isolated according to the manufacturer's protocols of QIAGEN midi prep. kit (QIAGEN Co., Germany) with an additional lysozyme treatment. The total 70 μg plasmid DNA of *B. thuringiensis mogi* was used to construct three libraries: (i) a GS FLX + shotgun library using the GS FLX + library preparation kit, (ii) an 8 kb-long paired-end library using the GS FLX paired-end kit, (iii) a HiSeq DNA shotgun library using the HiSeq2000 shot gun library kit. The libraries were sequenced using the Roche/454 pyrosequencing method on a Genome Sequencer FLX system (Macrogen, Korea) or the HiSeq™ 2000 platform (Illumina,San Diego, USA). In total, 64,395,859 and 40,850,371 bases were analyzed in single and paired-end reads, which yielded 10 million nucleotides

covering the genome ~16-fold (Table 5). Meanwhile, 1034 contigs were produced in 201 scaffolds through GS *de novo* assembler v2.6 (454 sequencing system software; Roche). Gaps within and between the scaffolds were confirmed and closed using primer walks and long-distance PCR amplification. End-sequencing of amplicons was carried out on an ABI 3730xl DNA Analyzer (Life Technologies). The complete sequences of chromosome (5,420,908 bp), pMOGI364 (364,564 bp) and pMOGI222 (222,348 bp) were determined.

2.3 Sequence annotation and analysis

Coding genes and pseudogenes across the genome were predicted using Glimmer (Delcher *et al.*, 1999), GeneMarkHMM (Lukashin and Borodovsky, 1998), and Prodigal (Hyatt *et al.*, 2010) and annotated by comparison with the NCBI-NR (Benson *et al.*, 2008). tRNA and rRNA were identified using tRNAscan-SE and RNAmmer, respectively. The annotation results were verified using Artemis (Rutherford *et al.*, 2000) and corrected manually gene by gene. Sequence similarities were determined using standalone BLAST programs (Altschul *et al.*, 1997) to search nucleotide and non-redundant protein databases from GenBank. Circular diagrams of plasmids were created using CGView server (http://stothard.afns.ualberta.ca/cgview_server/). Comparisons among related plasmids were made using BLAST programs

and multiple sequence alignments were performed using ClustalX (Thompson *et al.*, 1997).

Important *B. thuringiensis*-related sequences were also collected, including *cry*, δ endotoxin genes and mosquitocidal toxin for plasmid annotation and for surveying insecticidal genes.

3. RESULTS

3.1 General features of the genome sequence

The genome of *B. thuringiensis mogi* consists three replicons: a circular chromosome (5,420,908 bp) encoding 5,652 predicted open reading frames (ORFs), and two megaplasmids, pMOGI364 (364,564 bp), pMOGI222 (222,348 bp) (Table 6 and 7). The G+C content of the chromosome is 35.3%, while that of the plasmids are 31.3% and 34.2% , respectively. A total of 5,511 CDSs were identified in the chromosome. There are 102 tRNA genes representing all the 20 amino acids and 13 rRNA operons in the chromosome.

Comparison with *B. cereus* chromosomal maps suggests that all of these chromosomes have a similar organization in the half near the replication origin while displaying greater variability in the terminal half (Carlson *et al.*, 1996). The likely origin of replication of the chromosome of *B. thuringiensis mogi* was identified by

similarities to several features of the corresponding regions in *B. cereus* and other bacteria, including *dnaA* (chromosomal replication initiation protein, CDS g_0001mp) and *recF* (CDS g_0004mp) near the origin, GC nucleotide skew [(G-C)/(G+C)] analysis (Fig. 6), and the presence of multiple *dnaA* boxes and AT-rich sequences immediately upstream of the *dnaA* gene. The deduced replication termination site of the chromosome is believed to be localized near 2.6 megabases (Mb), according to GC skew analysis (Fig. 6), and the coding bias for the two strands of the genome is for the majority of CDSs to be on the outer strand from 0 to ~2.6 Mb and on the inner strand from ~2.6 Mb to the origin.

To reveal similarities and differences among the *B. cereus* group, genome of *B. thuringiensis mogi* was assembled in a pair-wise fashion. By running the MUMmer module according to the instructions of MUMmer 3.2, the results showed that *mogi* with two of the reference genomes, *B. anthracis* str. *Ames* (NC_003997.3) and *B. cereus* G9842 (NC_011772.1) (Fig. 7A and 7B) exhibit a much better synteny. While the other example, a comparison between *B. thuringiensis mogi* and *kurstaki* str. HD73 (NC_020238.1) seems to be difference, an inversion located on the blue segment of the region 2.7 Mb - 3.4 Mb.

Table 5. Read status of *B. thuringiensis* subsp. *mogi* genome assemblies.

No. of reads	No. of bases	Assembled	Partial	Singleton	Repeat	Outlier	Too short
309,746	100,868,054	296,142	6,229	4,936	2,003	436	0

- No. of reads: the read used in the assembly computation.
- No. of bases: the read's bases used in the assembly computation.
- Assembled: the read is fully incorporated into the assembly.
- Partial: only part of the read was included in the assembly.
- Singleton: the read did not overlap with any other reads in the input.
- Repeat: the read deemed to be from repeat regions.
- Outlier: the read was identified by the GS De Novo Assembler as problematic.
- Too short: the read was too short to be used in the computation.

Table 6. General characteristics of chromosome and gene prediction.

location	Size (bp)	G+C (%)	CDS	tRNA	5S rRNA	16S rRNA	23S rRNA
chromosome	5,420,908	35.3	5511	102	13	13	13

Table 7. General characteristics of two plasmids from strain *B. thuringiensis* subsp. *mogi.*

Plasmid	Size (bp)	G+C (%)	CDS	Pseudogene	Total gene	Coding density[a]
pMOGI364	364,564	31.3	357	53	410	1.125
pMOGI222	222,348	34.2	215	27	242	1.088

[a] The coding density is expressed in gene/kb.

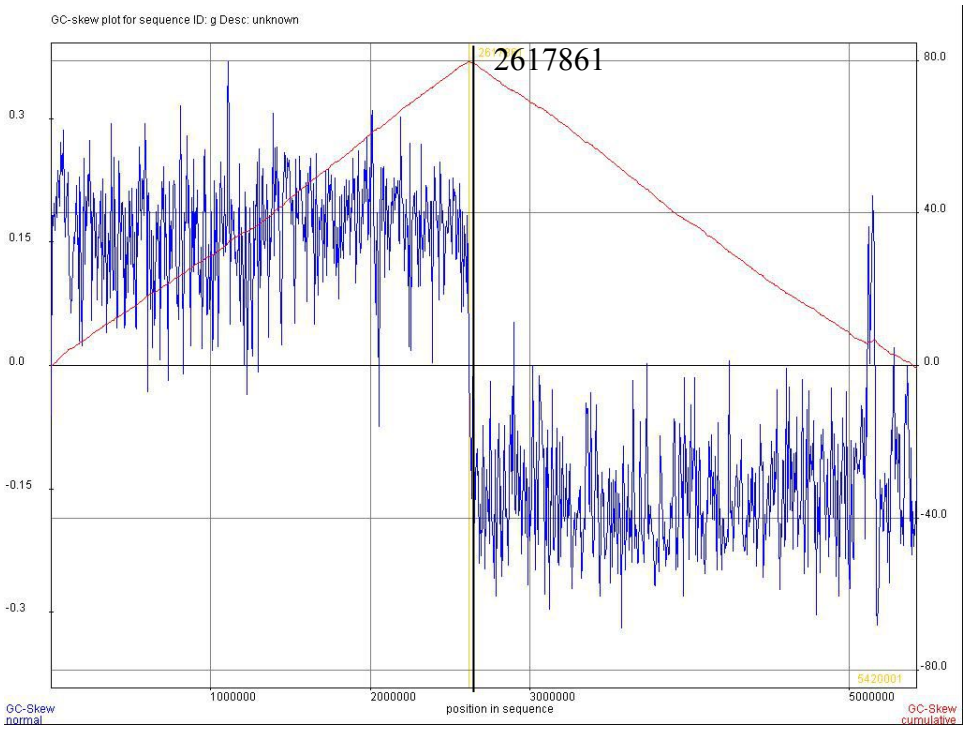


Fig. 6. GC-skew analysis of chromosome in *B. thuringiensis* subsp. *mogi*.

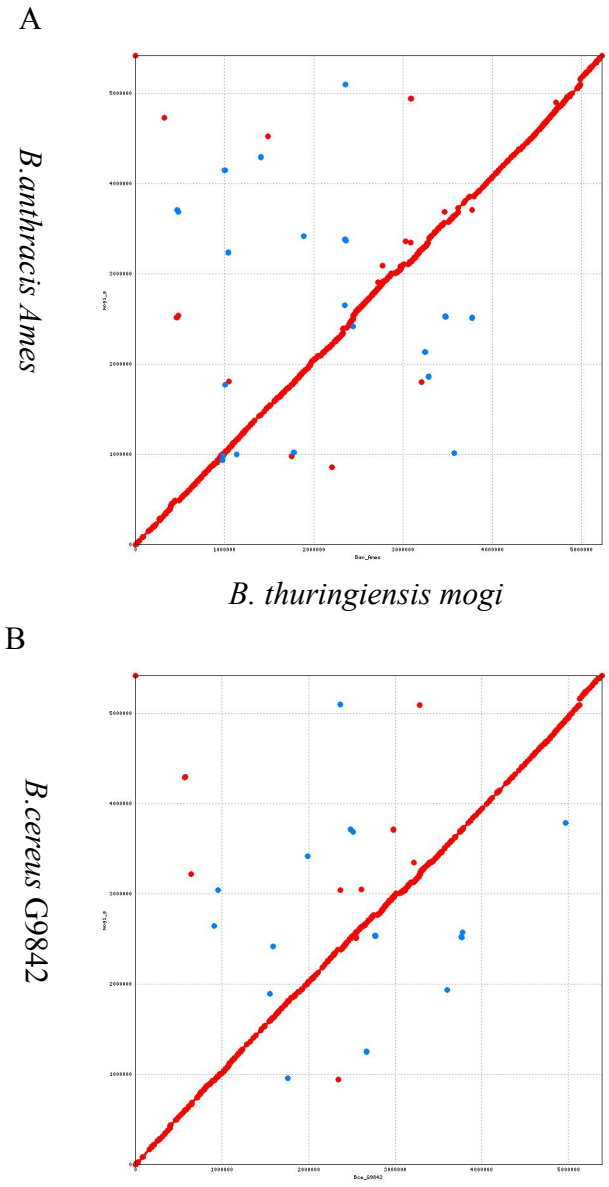
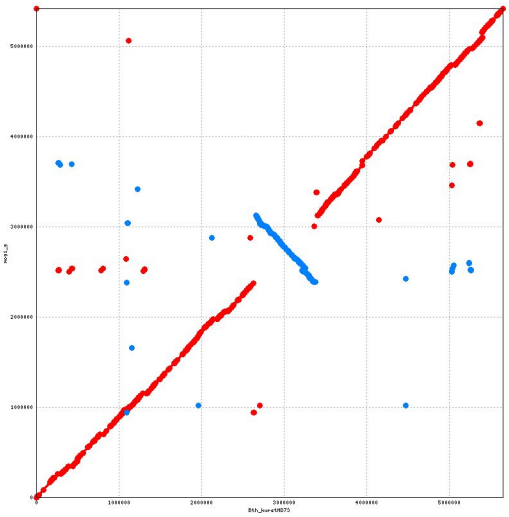


Fig. 7. Mummer analysis of *B. thuringiensis mogi* strain with the selected *Bacillus* species. (A) alignment of *B. thuringiensis mogi* to *B. anthracis Ames*; (B) alignment of *B. thuringiensis mogi* to *B. cereus* G9842. (C) alignment of *B. thuringiensis mogi* to *B. thuringiensis kurstaki* HD73. Matches in the forward strand are in red and those in the reverse strand are in blue. Linear regions are more conserved and scattered regions are less conserved.

C

B.thuringiensis kurstaki HD73



B. thuringiensis mogi

Fig. 7. Continued.

3.2 Phylogenic relationship between the *B. thuringiensis* subsp. *mogi* and the *B. cereus* group strains

The phylogenic tree was built based on 16S rRNA and DNA polymerase III α subunit (Zhao *et al*., 2007) sequences, using MEGA 5.2. Due to horizontal gene transfer, a major mechanism for genetic material exchange among prokaryotes, and the very close relations between members of *B. cereus* clade, which realized that phylogenic analysis may not yield stable relationship among the *B. cereus* group strains other than accurate and adequate selections of target sequences. Therefore, three sequences for this exercise: 16S rRNA, PolC, and DnaE, all of which are excellent for phylogeny were used. The latter two are DNA polymerase III α subunits of *Firmicute*s, whose sequences are highly conserved (Zhao *et al*., 2006; 2007). Although low bootstrap values indicated poor separations among the individual strains, the *B. thuringiensis mogi* strains showed much close to *B. cereus* in the 16S rRNA-based phylogeny (Fig. 8A). Furthermore, phylogeny based on PolC and DnaE (Fig. 8B) gave rise to a rather stable but fine differential relationship among all *B. cereus* group strains.

A

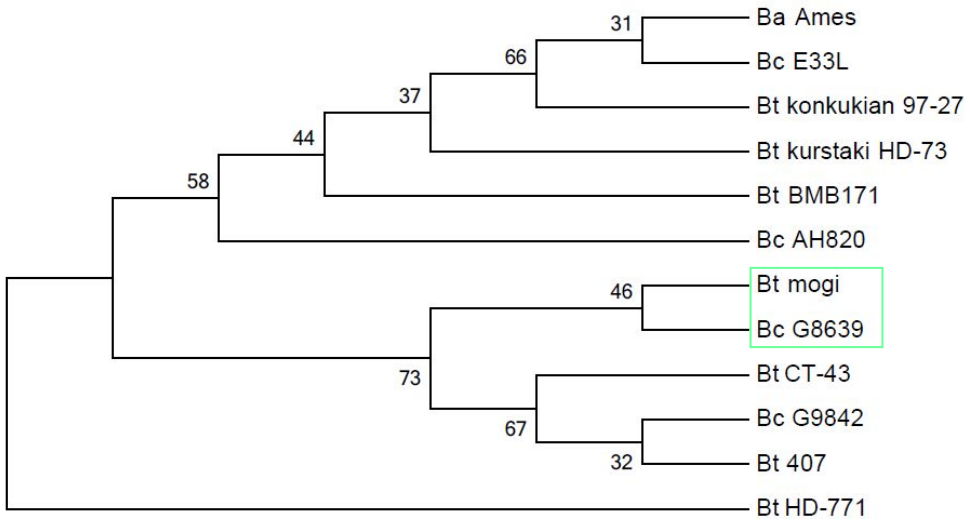


Fig. 8. Phylogenetic relationship of the *B. thuringiensis* subsp. *mogi* among the *Bacillus* clade. UPGMA trees were drawn with MEGA 5 following a sequential handling of selected CLUSTAL alignments coupled with bootstrapping. (A) Phylogenic survey based on 16S rRNA. (B) Phylogenic analysis of the isolates of *BacIllus* based on their PolC and DnaE genes.

B

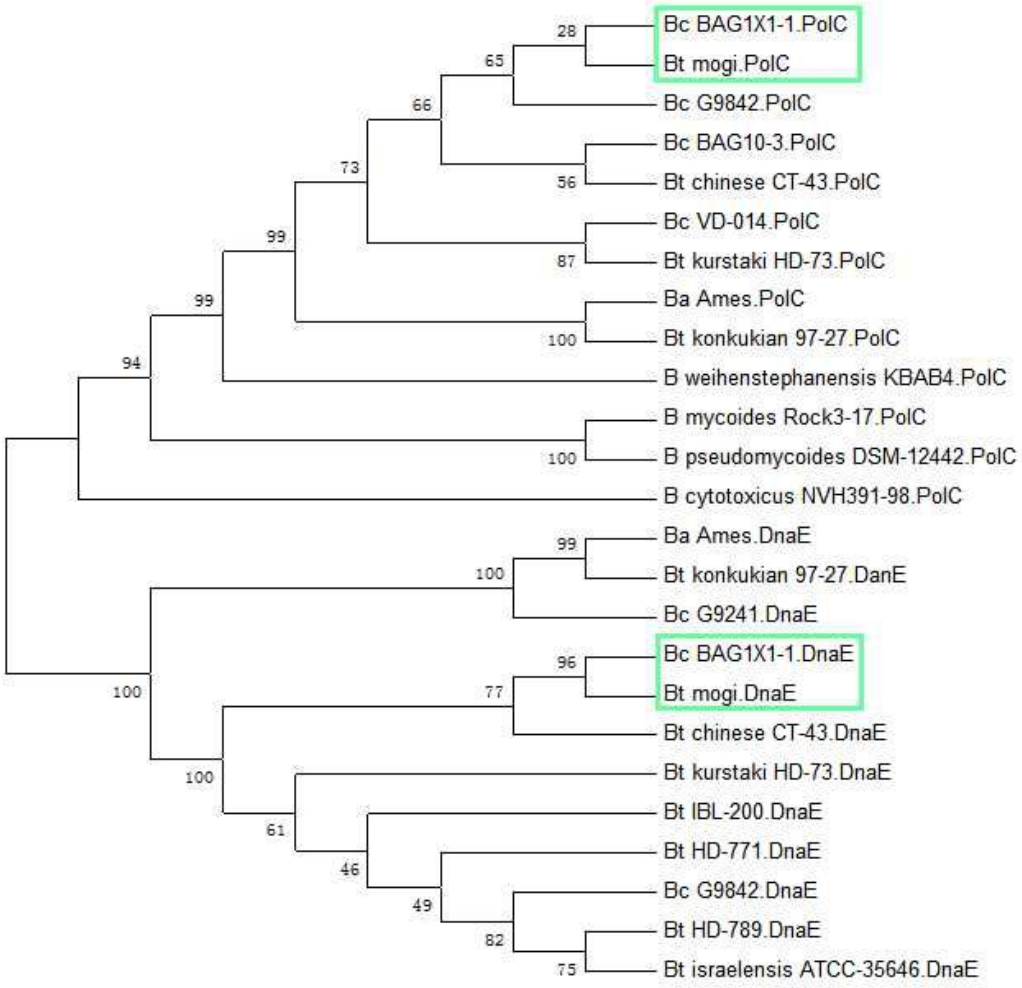


Fig. 8. Continued.

3.3 General features of plasmid sequence

Comparisons between the plasmid profiles of the novel serogroup *B. thuringiensis mogi* (H3a3b3d) and the H3 serotype (serovar *kurstaki*, *alesti*, *sumiyoshiensis* and *fukuokaensis*) showed that the novel serogroup isolate has a much simpler array of plasmids (Chapter 1, Fig. 1). This particular profile, and its lack of a relationship with type strains of the species, made the sequencing of extrachromosomal DNA present in *B. thuringiensis* subsp. *mogi* particularly interesting.

The complete nucleotide sequence of two plasmids from *B. thuringiensis* subsp. *mogi* was determined using two shotgun library plus 8 kb-long paired-end library sequencing method (see materials and method 2.2). The results and main characteristics of the plasmids are listed in Table 7. There are two megaplasmids, one approximately 364 kb, and one 222 kb plasmid. The plasmids were named pMOGI364 and pMOGI222 based on their sizes of 364,564 bp and 222,348 bp, respectively. All of them have G + C contents of around 30%, a value that is similar to that of plasmids from related species, such as *B. anthracis* and *B. cereus* (Andrup *et al*., 2003; Rasko *et al.*, 2005).

3.4 Plasmid pMOGI364

pMOGI364 properties are summarized in Table 7, and predicted genes are

described in Supplementary Table 1. A graphical representation of the plasmid is shown in Fig 11. pMOGI364 is the biggest plasmid from *B. thuringiensis* subsp. *mogi*. The complete DNA sequence was determined to be 364, 564 bp with a G+C content of 31.3%, within the range characteristic for the *B. thuringiensis* species. Of the 410 putative genes, 172 (42%) could be assigned putative functions, 186 (45.4%) encoded conserved hypothetical proteins, and 50 (12.2%) were pseudogenes.

(i) Analysis of the pMOGI364 sequence for the identification of the putative replicon

Most plasmids with sequenced genomes contain genes encoding a replication initiator protein. An alignment of such genes has shown that various plasmids can be grouped into plasmid families which share significant homologies in their replication initiator genes and the origin of replication (Chattoraj, 2000; del Solar *et al*., 1998; Khan, 1997). The genes encoding replication initiator proteins of both chromosome and plasmids are generally located in the vicinity of their replication origins. However, in the homology searches using a variety of software was failed to identify homologs of known replication initiator proteins in the pMOGI364 sequence. Then GC skew (Fig. 12), strand-specific biases such as gene orientation, plasmid-specific oligomer skew analysis, and origin comparisons (provided by the Genome Atlas Database at http://www.cbs.dtu.dk/services/GenomeAtlas/) were used to predict the location of an origin of replication (*ori*). The result suggested the possible location of the putative

ori of pMOGI364 was near nt 1 and the termination site was near nt 130,000.

(ii) FtsZ/tublin-related protein in pMOGI364

CDS pMOGI364_403, which shows high homology (99% Max identity) with FtsZ/tubulin-related protein, could potentially function in plasmid partitioning. The prokaryotic FtsZ protein is a polymer-forming GTPase that shares structural and functional similarities with eukaryotic tubulins (Vaughan *et al*., 2004). FtsZ assembles into a ring structure on the inner surface of the cytoplasmic membrane at the site of cell division. The so-called Z ring is progressively reduced in diameter, a process which leads to invagination of the dividing septum. Meanwhile, the CDS also shares limited homology (26% identity) with that encodes by pXO1-45 (*repX*), which was required for the replication of the miniplasmid in *B. anthracis* (Tinsley and Khan, 2006). In addition, ORF pBt156, which encodes a peptide with weak amino acid similarity to the FtsZ/tubulin-like proteins of *Pyrococcus* (BAB17294) and pXO1-45, are essential for replication of pBtoxis (Tang *et al*., 2006).

(iii) DNA topoisomerase III in *mogi*

Interestingly, pMOGI364 show similarity to *B. cereus* G9842 pG9842_209. About 200 kb sequence from pMOGI364, show a very high similarity (more than 90% identity) with the plasmid pG9842_209 (Fig. 13 and 14), and the last 146 kb fragment is found to harbor several *cry* genes. The level of protein similarity, combined with the conservation of gene order, suggests that these plasmids might

have diverged recently. pMOGI364 and pG9842_209 share a common backbone (Fig. 13). The insertion sites on pMOGI364 are between CDS pMOGI364_207 and pMOGI364_360, two CDS which combine together code for a intact type I DNA topoisomerase.

DNA topoisomerases resolve entangled DNA intermediates by transiently cleaving one or two DNA strands and passing another intact strand(s) through the nick. Based on their catalytic mechanism, topoisomerases have been categorized into four subfamilies, type IA and IB, and type IIA and IIB. Type IA topoisomerases are highly conserved from bacteria to humans. In most cells, two type IA topoisomerases are present while among the *Bacillus, B. anthracis* (Read *et al*., 2003), *B. cereus* (Ivanova *et al*., 2003) and *B. thuringiensis* have three chromosomal copies of type IA topoisomerases. Two plasmids, pXO1 and pXO2 which can be mobilized in *B. anthracis,* appear to encode two additional type IA topoisomerases. Therefore, five type IA topoisomerases may be present in these bacteria.

Here, DNA topoisomerase III from *B. thuringiensis* subsp. *mogi* are summarized in Table 8, there are five topoisomerase III on chromosomal and three on plasmid pMOGI364. Among them, four topoisomerase (g_0397mp, g_0413mp, g_1847mp, g_1858mp) from chromosomal are truncated and only one topoisomerase (g_0404mp) is intact. The same as topoisomerase on pMOGI364, two of them (pMOGI364_207, pMOGI364_360) are truncated and one (pMOGI364_233) is intact (Fig. 13). The

megalign results showed that the two neighbored truncated topoisomerase join together could be an intact topoisomerase III (g_0397mp plus g_0413mp, Fig 14-1; g_1847mp plus g_1858mp, Fig. 14-2; pMOGI364_207 plus pMOGI364_360, Fig. 14-3).

Prokaryotic type IA topoisomerases may have more diverse functions than simply the maintenance of genomic stability. These enzymes may have roles in the horizontal gene transfer of promiscuous plasmids or conjugational transposons. Type IA topoisomerases may also function in the process of conjugational DNA transfer, transposon integration (Sutanto *et al*., 2002), plasmid maintenance and plasmid segregation. To deal with stringent environmental stressors such as UV, chemical and free radical damage, bacteria may not only need sophisticated DNA damage repair systems, but also efficient DNA recombination systems to create, adopt and spread endogenous or exogenous mutations. Therefore, different type IA topoisomerases may be required in a variety of DNA repair and recombination processes. In addition, the characteristic cellular development stages such as sporulation and germination of spore-forming Gram-positive bacteria may require DNA replication and recombination systems as well (Li *et al*., 2005). These results predicted that the intact topoisomerase III on pMOGI364 may be essential for this strain.

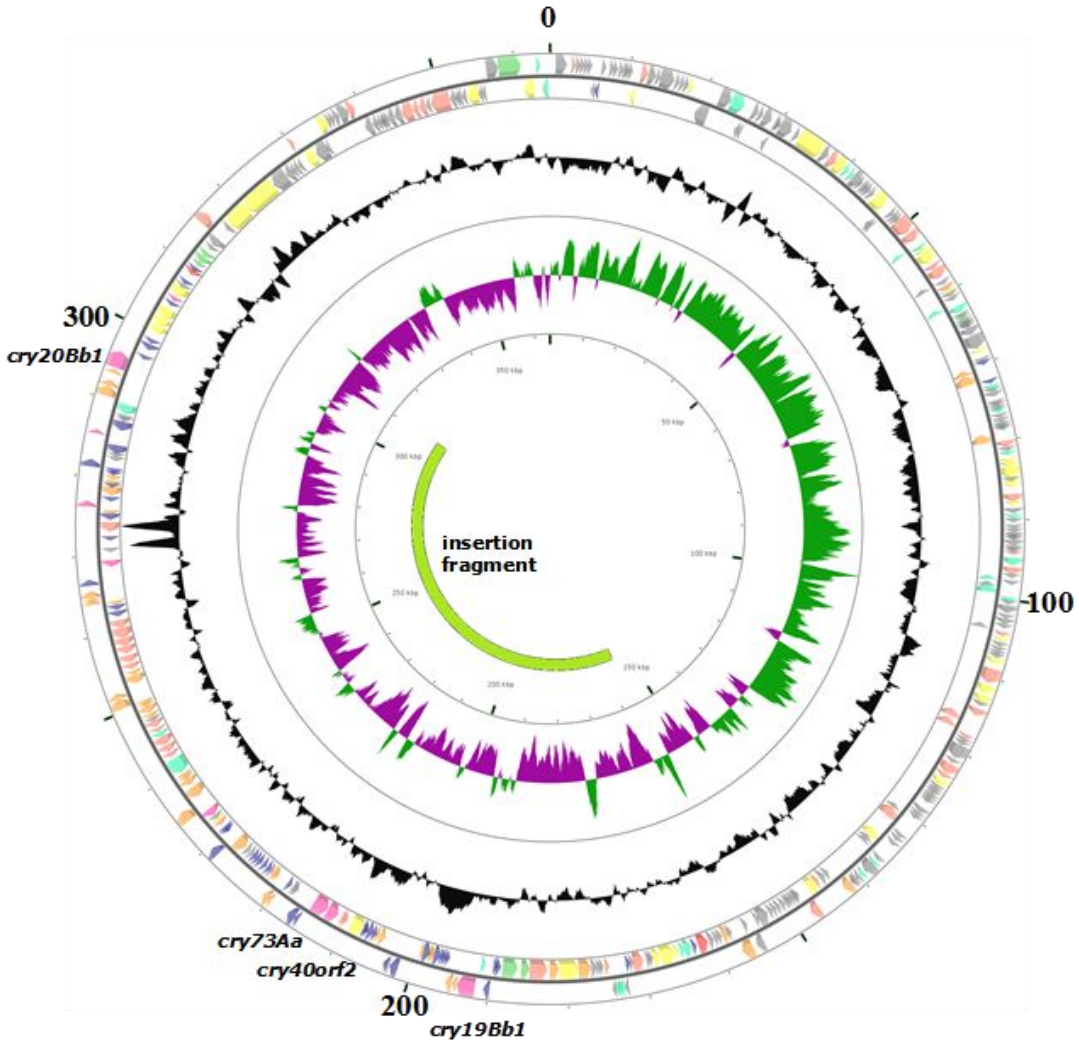


Fig. 9. Circular representation of pMOGI364. The inner circle represents GC bias

[(G- C)/(G+C)], with positive values in green and negative values in purple; the second circle represents GC content; and the outer two circles represent predicted genes on the reverse and forward strands (selected toxin-related CDSs are marked for reference). Predicted function/homologies are indicated by the color key featured below. The outer scale is marked in kilobases. The CDS number corresponds to Supplementary Table 1.

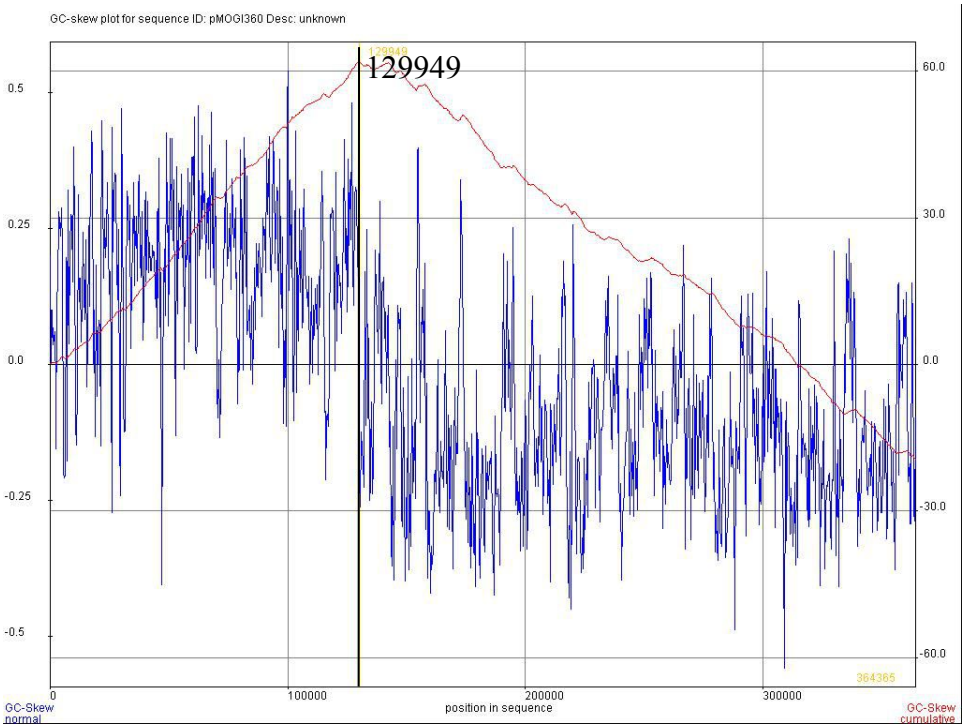


Fig. 10. GC-skew of the plasmid pMOGI364 in *B. thuringiensis* subsp. *mogi*.

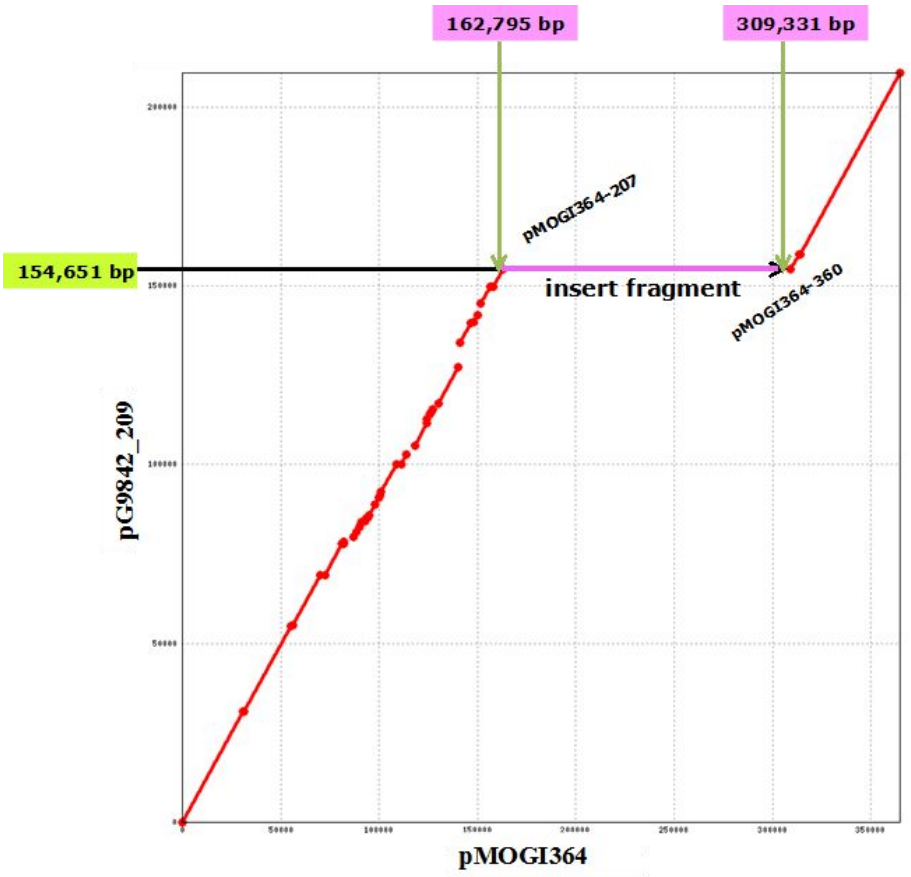
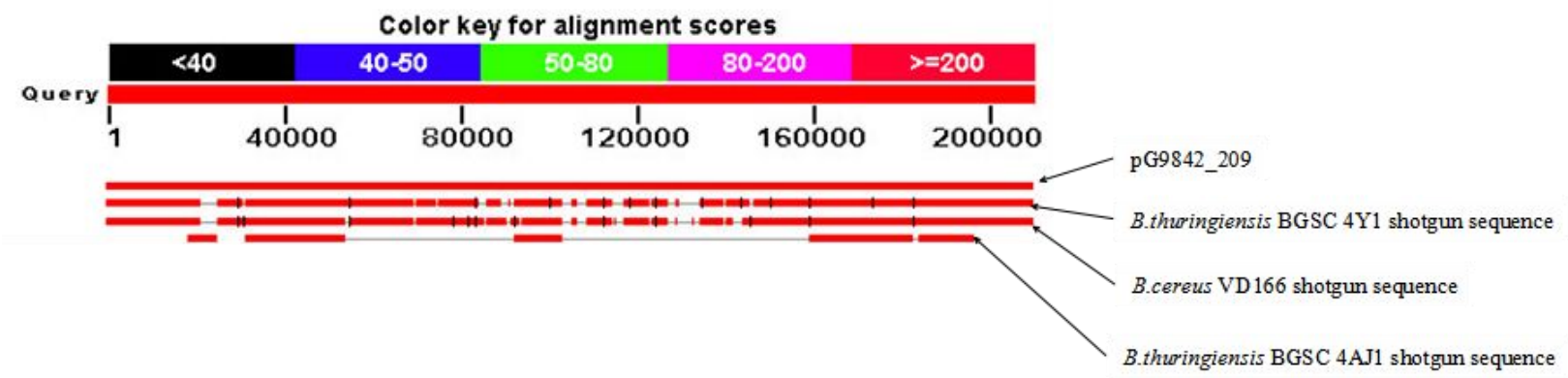


Fig. 11. Mummer analysis of pMOGI364 with pG9842_209.

(A)



(B)

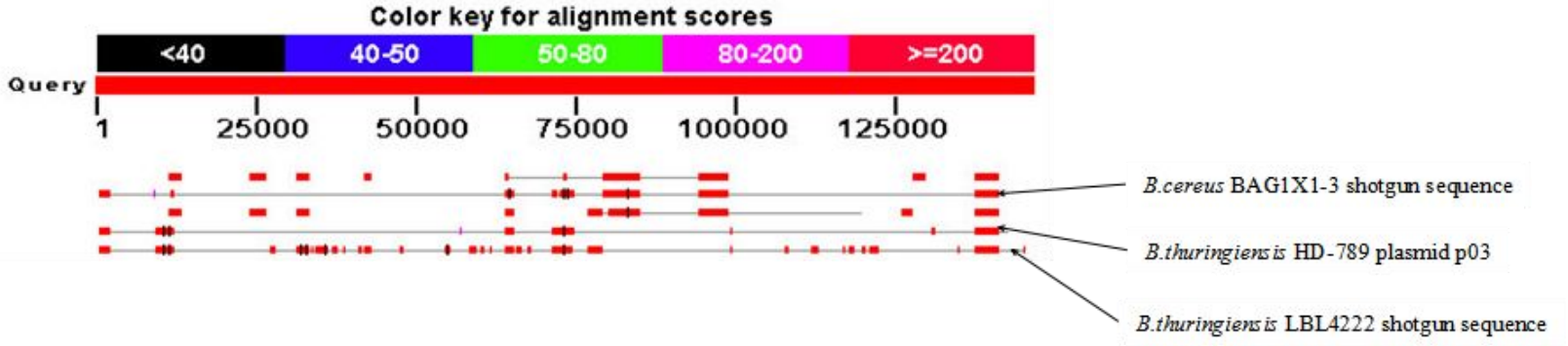


Fig. 12. Graphic overview of the results of megablast search using the sequence of pMOGI364 218 kb fragment (A) and 146kb insertion fragment (B) as a query against the database.

Table 8. DNA topoisomerase III in *B. thuringiensis* subsp. *mogi*.

No.	ltag	size	strand	predicted product homology with	(% aa identity)
1-1	g_0397mp	562 aa	-	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002443902.1) (729 aa)	98% in 543 aa
1-2	g_0413mp	198 aa	-	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002443902.1) (729 aa)	99% in 192 aa
2	g_0404mp	729 aa	+	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002443902.1) (729 aa)	90% in 730 aa
3-1	g_1847mp	369 aa	+	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002445288.1) (714 aa)	99% in 367 aa
3-2	g_1858mp	348 aa	+	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002445288.1) (714 aa)	99% in 348 aa
4-1	pMOGI360_207	446 aa	-	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002454832.1) (687 aa)	97% in 445 aa
4-2	pMOGI360_360	253 aa	-	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002454832.1) (687 aa)	96% in 243 aa
5	pMOGI360_233	716 aa	-	DNA topoisomerase III [*Bacillus cereus* HuB4-4] (EOP91371.1) (716 aa)	95% in 716 aa

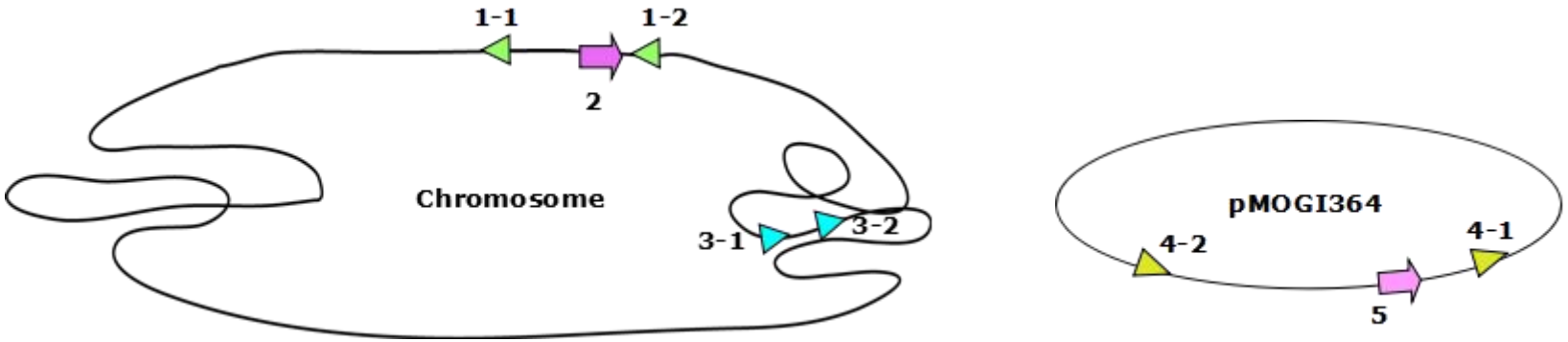


Fig. 13. Previews about DNA topoisomerase III in *B. thuringiensis* subsp. *mogi.* Truncated DNA topoisomerase III genes are showed in triangles and the intact DNA topoisomerase III genes are mapped in arrows.

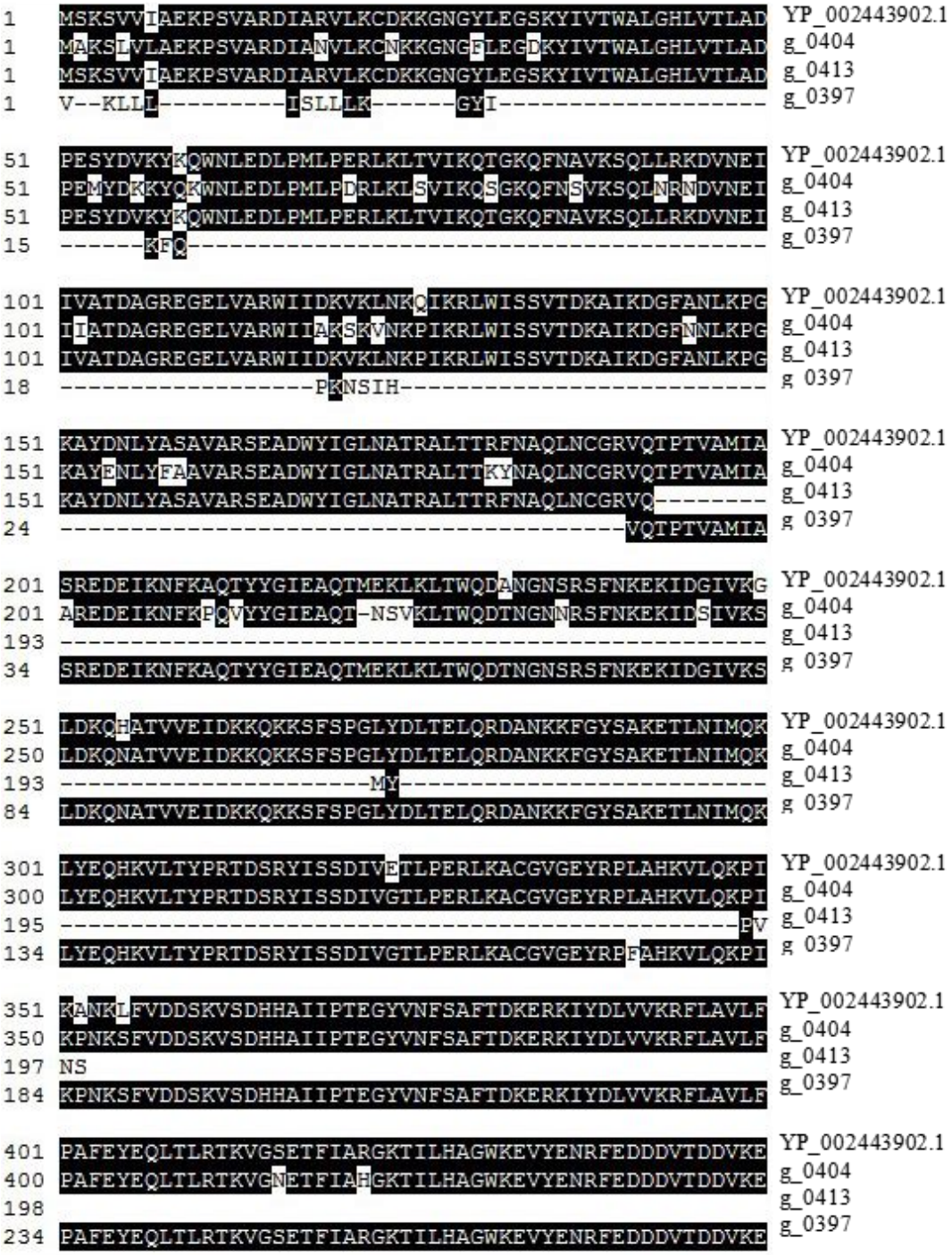
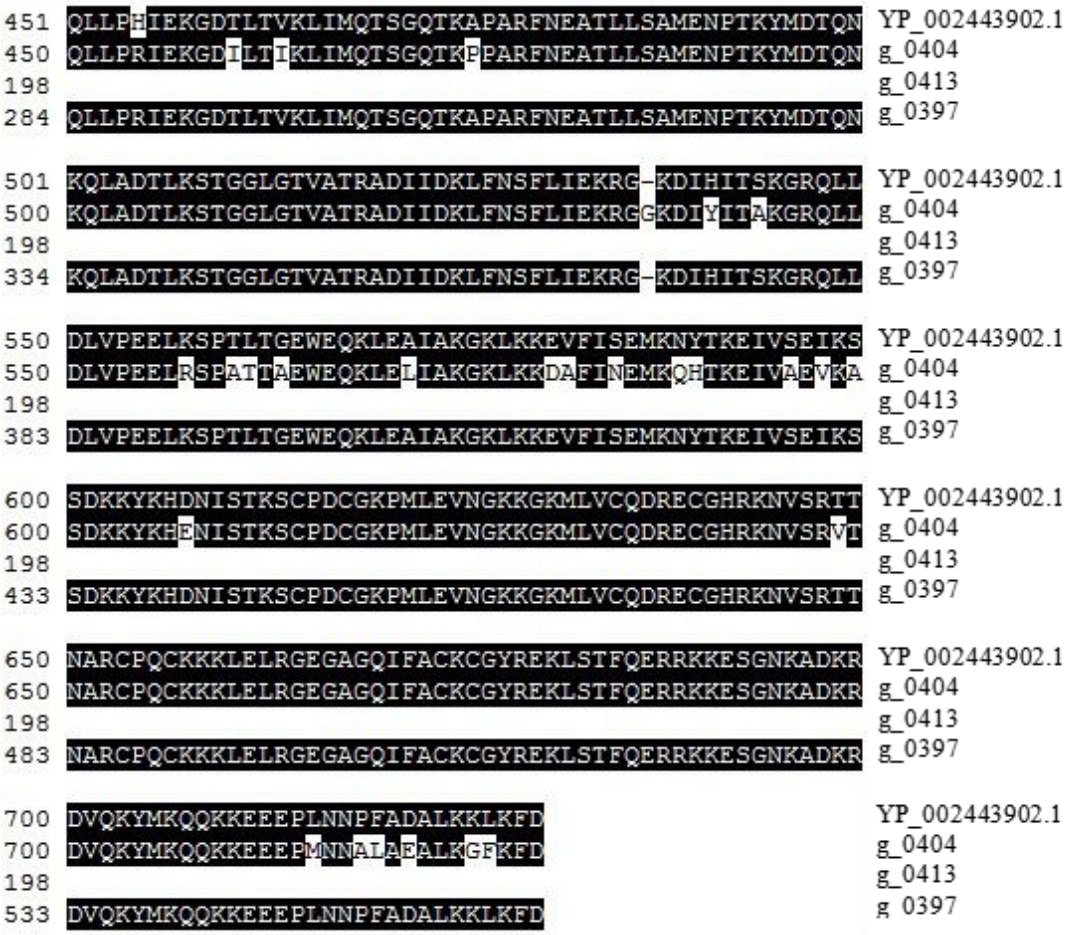


Fig. 14-1. Sequence alignment analysis of the truncated DNA topoisomerase with the intact one from NCBI database. Protein sequence comparisons were performed using the MegAlign program.

- continued


451 QLLPHIEKGDTLTVKLIMQTSGQTKAPARFNEATLLSAMENPTKYMDTQN YP_002443902.1
450 QLLPRIEKGDILTIKLIMQTSGQTKPPARFNEATLLSAMENPTKYMDTQN g_0404
198 g_0413
284 QLLPRIEKGDTLTVKLIMQTSGQTKAPARFNEATLLSAMENPTKYMDTQN g_0397
501 KQLADTLKSTGGLGTVATRADIIDKLFNSFLIEKRG-KDIHITSKGRQLL YP_002443902.1
500 KQLADTLKSTGGLGTVATRADIIDKLFNSFLIEKRGGKDIYITAKGRQLL g_0404
198 g_0413
334 KQLADTLKSTGGLGTVATRADIIDKLFNSFLIEKRG-KDIHITSKGRQLL g_0397
550 DLVPEELKSPTLTGEWEQKLEAIAKGKLKKEVFISEMKNYTKEIVSEIKS YP_002443902.1
550 DLVPEELRSPATTAEWEQKLELIAKGKLKKDAFINEMKQHTKEIVAEVKA g_0404
198 g_0413
383 DLVPEELKSPTLTGEWEQKLEAIAKGKLKKEVFISEMKNYTKEIVSEIKS g_0397
600 SDKKYKHDNISTKSCPDCGKPMLEVNGKKGKMLVCQDRECGHRKNVSRTT YP_002443902.1
600 SDKKYKHENISTKSCPDCGKPMLEVNGKKGKMLVCQDRECGHRKNVSRVT g_0404
198 g_0413
433 SDKKYKHDNISTKSCPDCGKPMLEVNGKKGKMLVCQDRECGHRKNVSRTT g_0397
650 NARCPQCKKKLELRGEGAGQIFACKCGYREKLSTFQERRKKESGNKADKR YP_002443902.1
650 NARCPQCKKKLELRGEGAGQIFACKCGYREKLSTFQERRKKESGNKADKR g_0404
198 g_0413
483 NARCPQCKKKLELRGEGAGQIFACKCGYREKLSTFQERRKKESGNKADKR g_0397
700 DVQKYMKQQKKEEEPLNNPFADALKKLKFD YP_002443902.1
700 DVQKYMKQQKKEEEPMNNALAEALKGFKFD g_0404
198 g_0413
533 DVQKYMKQQKKEEEPLNNPFADALKKLKFD g_0397

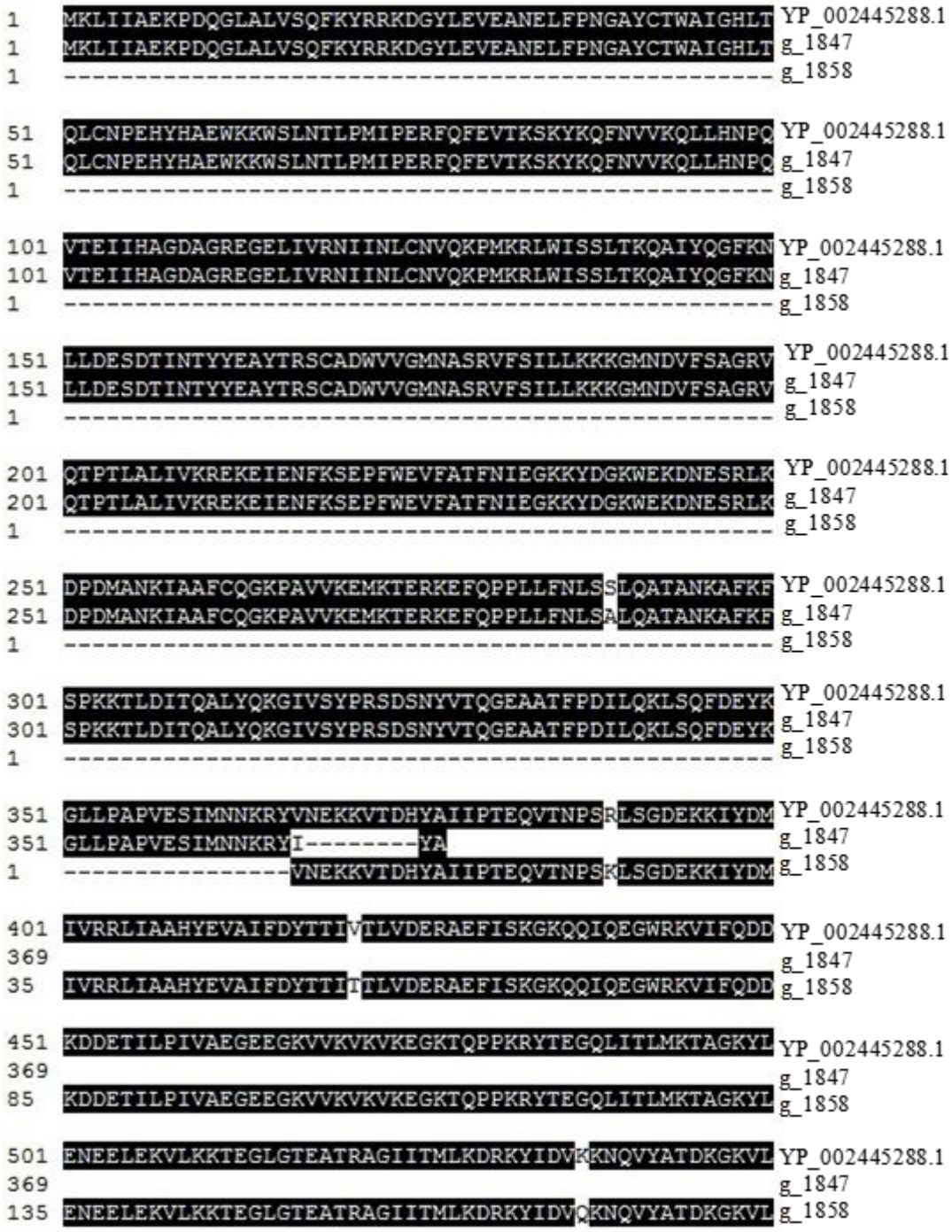


Fig. 14-2. Sequence alignment analysis of the truncated DNA topoisomerase with the intact one from NCBI database. Protein sequence comparisons were performed using the MegAlign program.

- Continued.


551 ITAIGDKILASPEMTAKWEQRLAEIGEGTASPATFMEQTKKLSAKIIEDA YP_002445288.1
369 g_1847
185 ITAIGDKILASPEMTAKWEQRLAEIGEGTASPATFMEQTKKLSAKIIEDA g_1858
601 VEMSEKWDFTGLHVESIERKGSKFTTGKKVGSCKKCDGDVIDKSTFYGCS YP_002445288.1
369 g_1847
235 VEMSEKWDFTGLHVESIERKGSKFTTGKKVGNCKKCDGDVIDKSTFYGCS g_1858
651 NYNTTQCDFTISKKILSKTISQKNMTKLLKGEKTDLIKGFKKGEKTFDAK YP_002445288.1
369 g_1847
285 NYNTTQCDFTISKKILSKTISQKNMTKLLKDEKTDLIKGFKKGEKTFDAK g_1858
701 LEWKDNKINFVFEN YP_002445288.1
369 g_1847
335 LEWKDNKINFVFEN g_1858

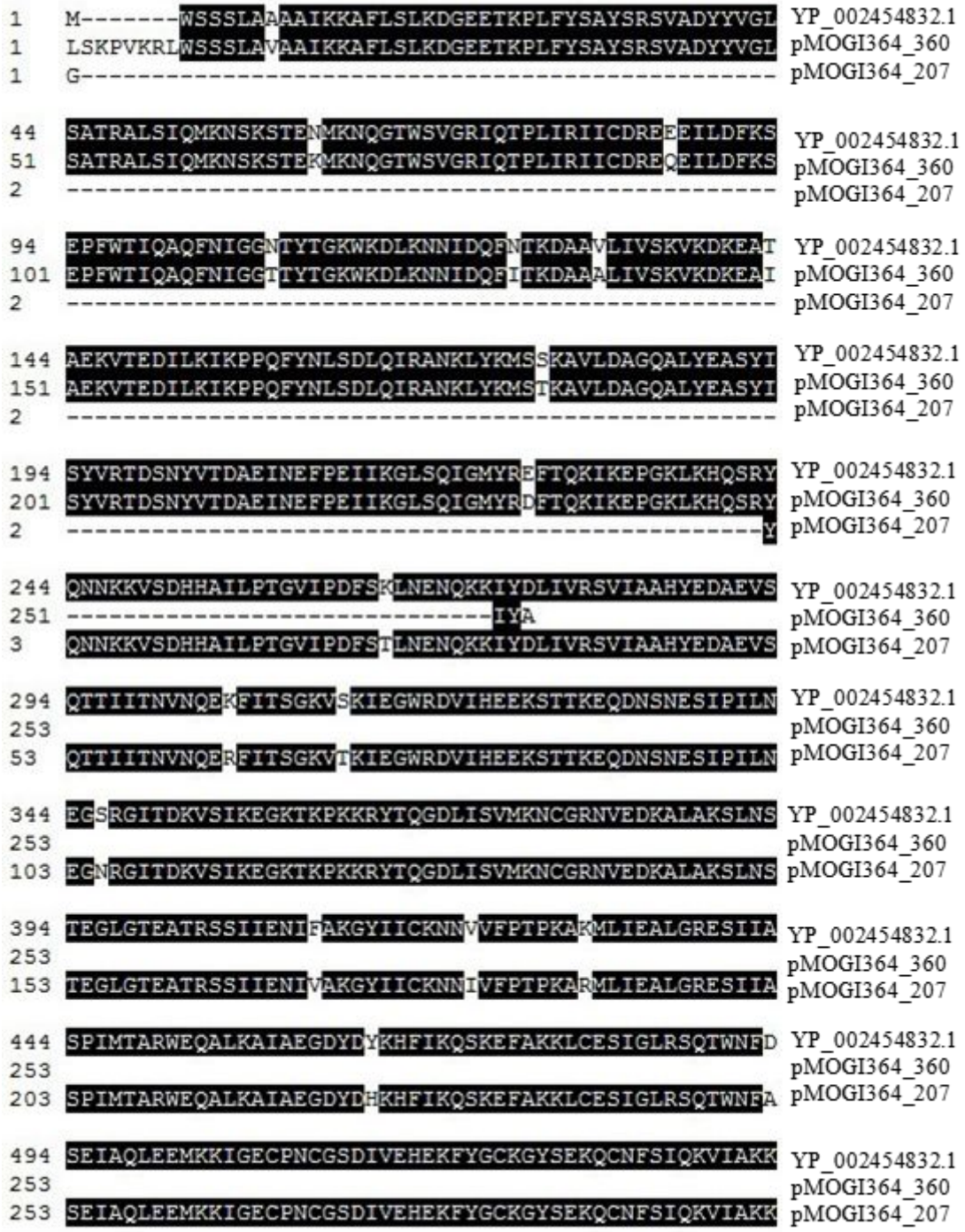


Fig. 14-3. Sequence alignment analysis of the truncated DNA topoisomerase with the intact one from NCBI database. Protein sequence comparisons were performed using the MegAlign program.

- continued


544 KISPAQVKKLLKDKKTDVIKGFKSSKSERTFETFLYYDSEKKCIDWGFNQ YP_002454832.1
253 pMOGI364_360
303 KISPAQVKKLLKDKKTDVIKGFKSSKSERTFETFLYYDSEKKCIDWGFNQ pMOGI364_207
594 AKNDNKPSSKDTGFKCPLCKDNLVEHQKFIGCSGFKNGCEFKISKNICGV YP_002454832.1
253 pMOGI364_360
353 AKNDNKPSSKDTGFKCPLCKNNLVEHQKFIGCSGYKNGCEFKISKNICGV pMOGI364_207
644 NLTPTHIEELVNNGETSMIDGFVFKDKTFRKALCVIDGKVMFKK YP_002454832.1
253 pMOGI364_360
403 NLTSTHIEELVNNGETSMIDGFVFKDNTFRKALCVIDGKVMFKK pMOGI364_207

3.5 Plasmid pMOGI222

The 222,348 bp sequence of plasmid pMOGI222 has 34.2% G + C content and 242 predicted genes. Twenty-seven (11.2%) of the predicted genes are pseudogene and 90 (37.2%) are hypothetical, whereas 7 CDS showed homology with toxin-related genes in NCBI database sequences, and 1 gene was thought to enhance crystal formation and subsequent cell viability by acting as chaperon (Supplementary Table 2). A graphical representation of the plasmid is shown in Fig. 15.

The protein proposed as responsible for the replication of pMOGI222 (Rep, replication initiation protein, pMOGI222_003) showed a high similarity to the Rep proteins of the pAMβ1 family of theta-replicating plasmids, such as Rep165 of pBMB165 (96.5% identity), Rep of pBT9727 (95.8% identity), Rep63A of pAW63 (82.5% identity) and RepS of pXO2 (81.1% identity) from *B. anthracis* (Tinsley *et al*., 2004), RepE of pAMβ1 (42.1% identity) from *Enterococcus faecalis* (Bruand *et al*., 1993) (Fig. 18). These conjugative plasmids replicate by a theta-type mechanism, and their replication proceeds unidirectionally from the origin (Bruand *et al*., 1993).

In addition, two cis-functioning regions were found near the Rep protein in pMOGI222. One cis-functioning region, located immediately downstream of Rep, displayed significant similarity to the cis-functioning origin of replication (*ori*) harbored in the corresponding locus of the pAMβ1 family plasmids cited above (Fig. 19). The other cis-functioning region was constituted of iterons (Fig. 20). The iteron

region was located between the copy number control gene (*parA*, pMOGI222_002) and the replication gene (*rep*, pMOGI222_003). These were sets of repeated DNA sequences that have been reported to serve as a binding site for the replication initiation protein and thus to play an important role in the replication and/or the control of replication in other theta-replicating plasmids. In addition, a AT-rich DNA region in the vicinity of the iteron sequences was found and marked with underline in Fig. 20.

Taken together, the conservation among these Rep proteins and *ori* regions thus provided significant evidence that pMOGI222 belongs to the pAMβ1 family of Gram-positive theta-replicating plasmids. All theta replication proteins from *B.cereus* group plasmids share a significant level of similarity and cluster together phylogenetically suggesting a common ancestral origin (Fig. 21).

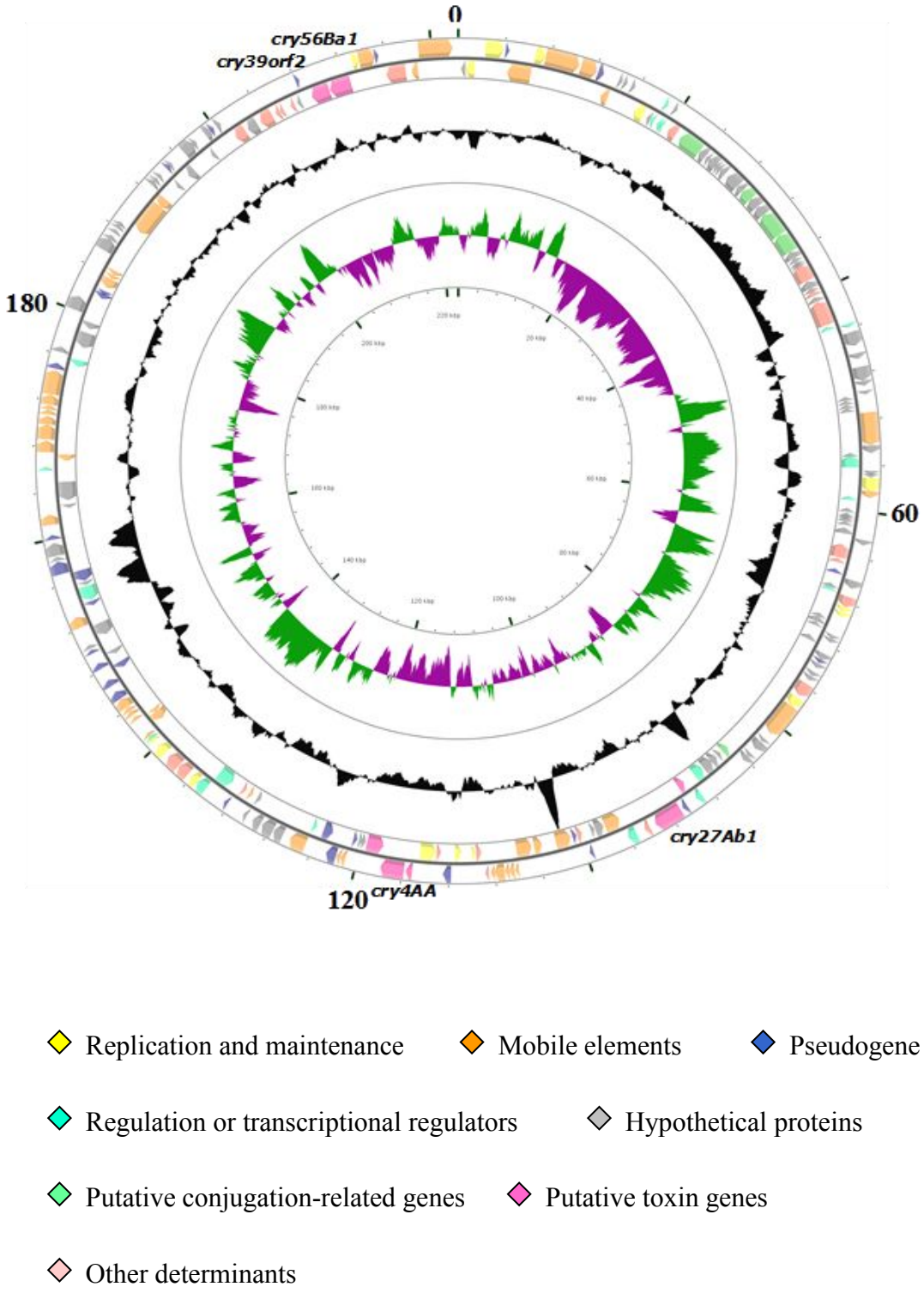


Fig. 15. Circular representation of pMOGI222. The inner circle represents GC bias

[(G- C)/(G+C)], with positive values in green and negative values in purple; the second circle represents GC content; and the outer two circles represent predicted genes on the reverse and forward strands (selected toxin-related CDSs are marked for reference). Predicted function/homologies are indicated by the color key featured below. The outer scale is marked in kilobases. The CDS number corresponds to Supplementary Table 2

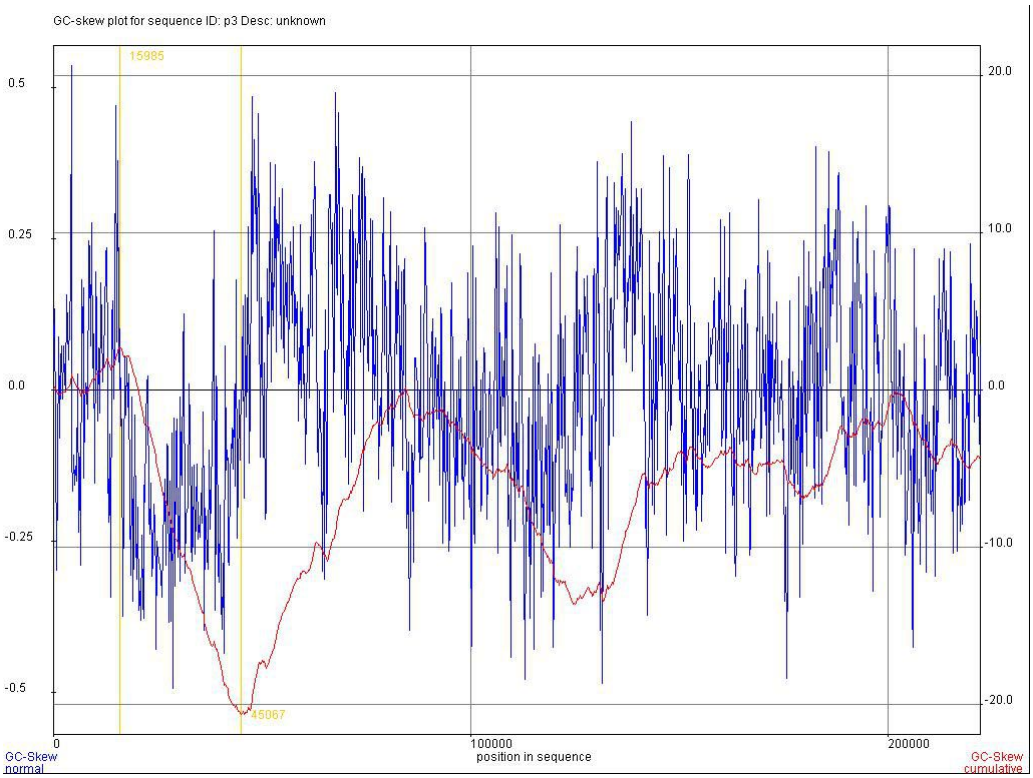


Fig. 16. GC-skew of the plasmid pMOGI222 in *B. thuringiensis* subsp.*mogi.*

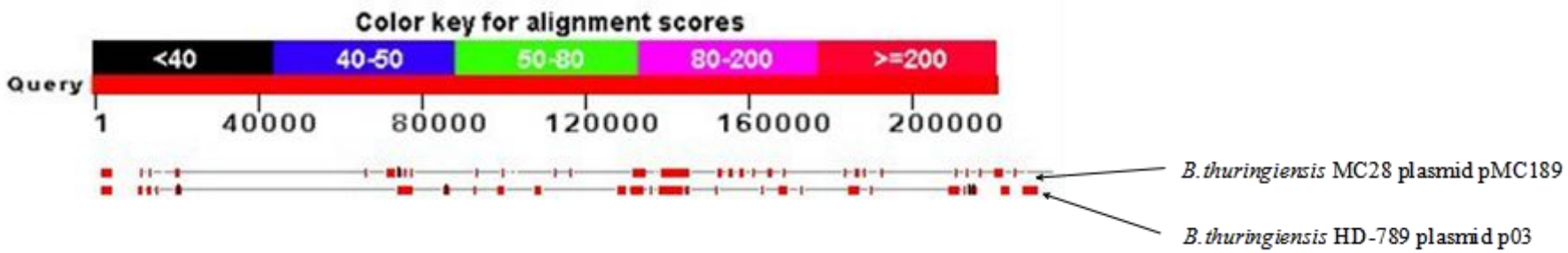
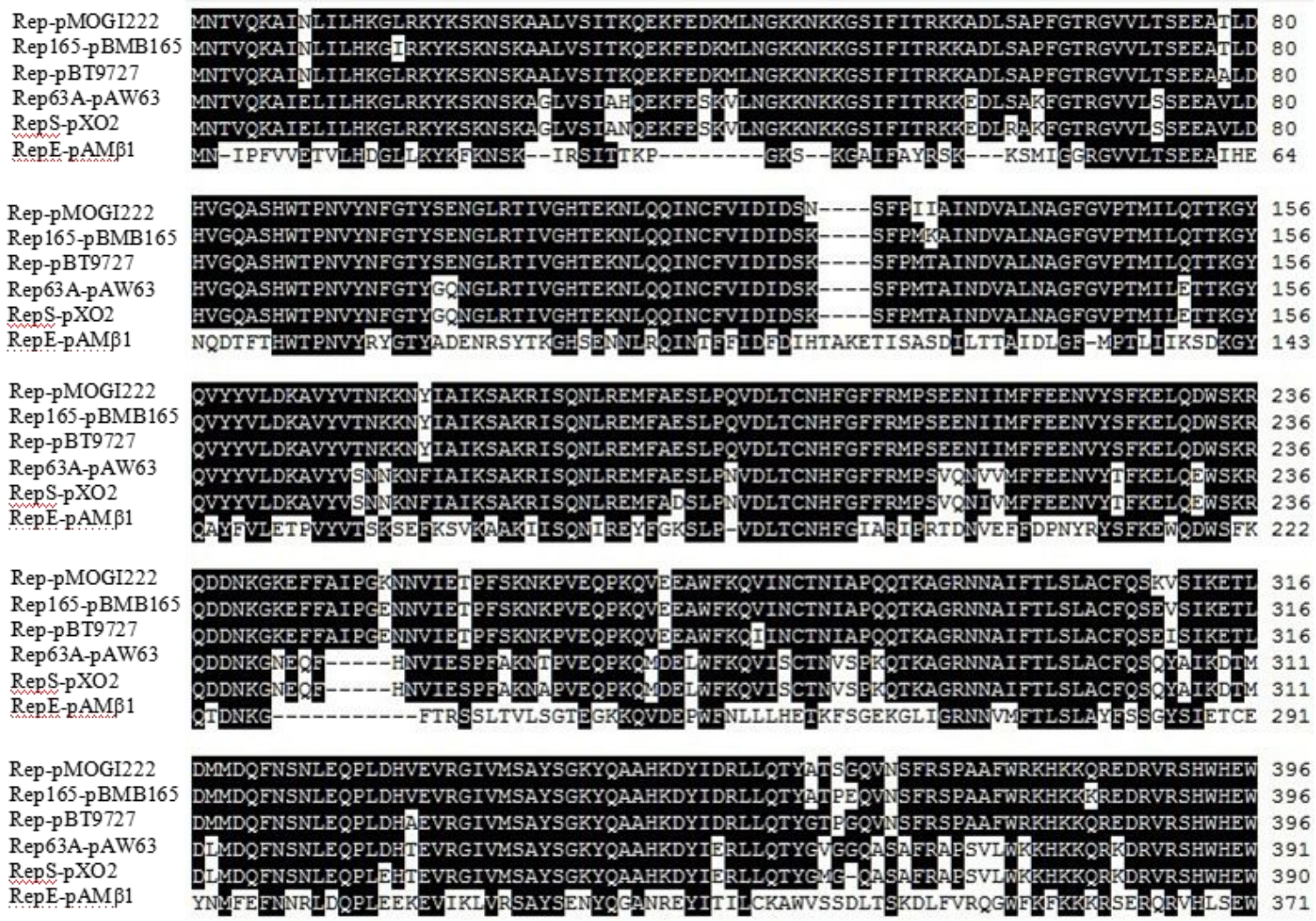


Fig. 17. Graphic overview of the result of a megablast search using the sequence of pMOGI222 as a query against the database.


Rep-pMOGI222
Rep165-pBMB165
Rep-pBT9727
Rep63A-pAW63
RepS-pXO2
RepE-pAMβ1

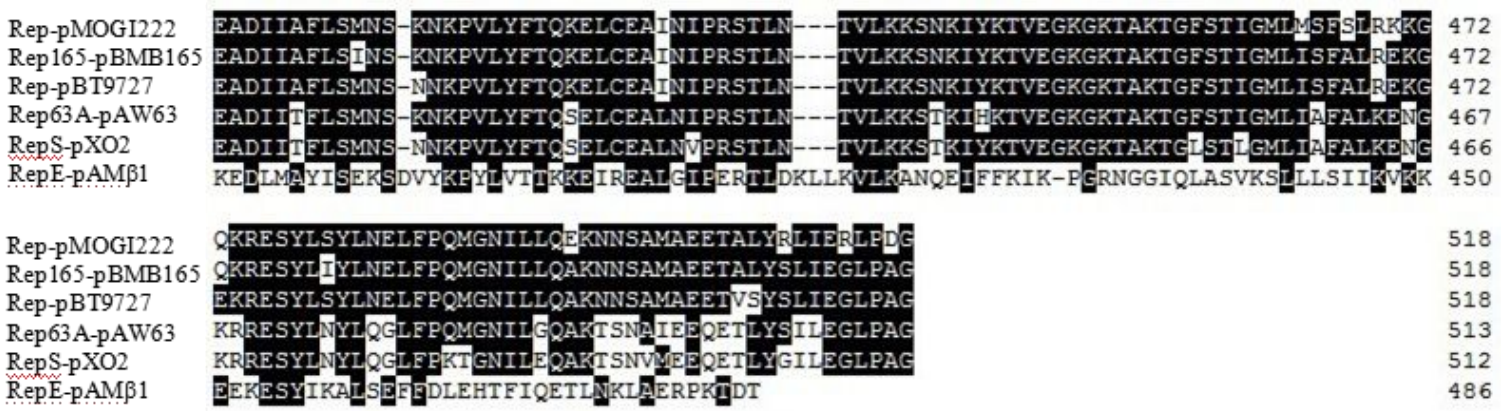


Fig. 18. Alignment of the predict Rep protein of pMOGI222, Rep165 protein of plasmid pBMB165, Rep63A protein of plasmid pAW63, RepS protein of pXO2, and RepE protein of pAMβ1 of these three plasmids. The alignment was done using the ClustalW program, and the shaded letters indicate identical amino acids.

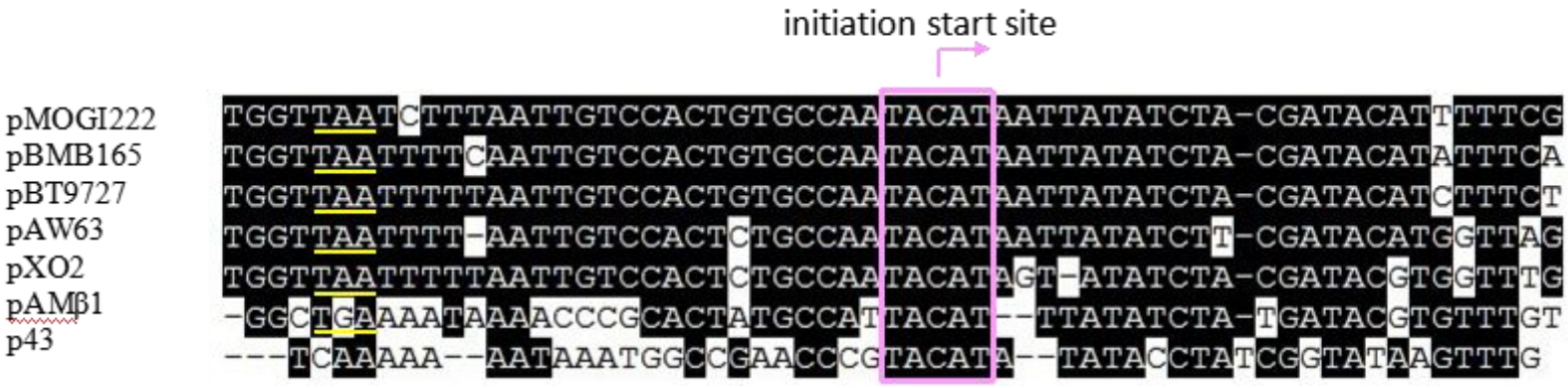


Fig. 19. Alignment of the origin of replication in pMOGI222, pBMB165, pBT9727, pAW63, pXO2, pAMβ1 and p43. The shaded letters indicate identical nucleotides. The box highlights a conserved replication initiation site 'TACAT'. The stop codons of the *rep* genes are underlined.

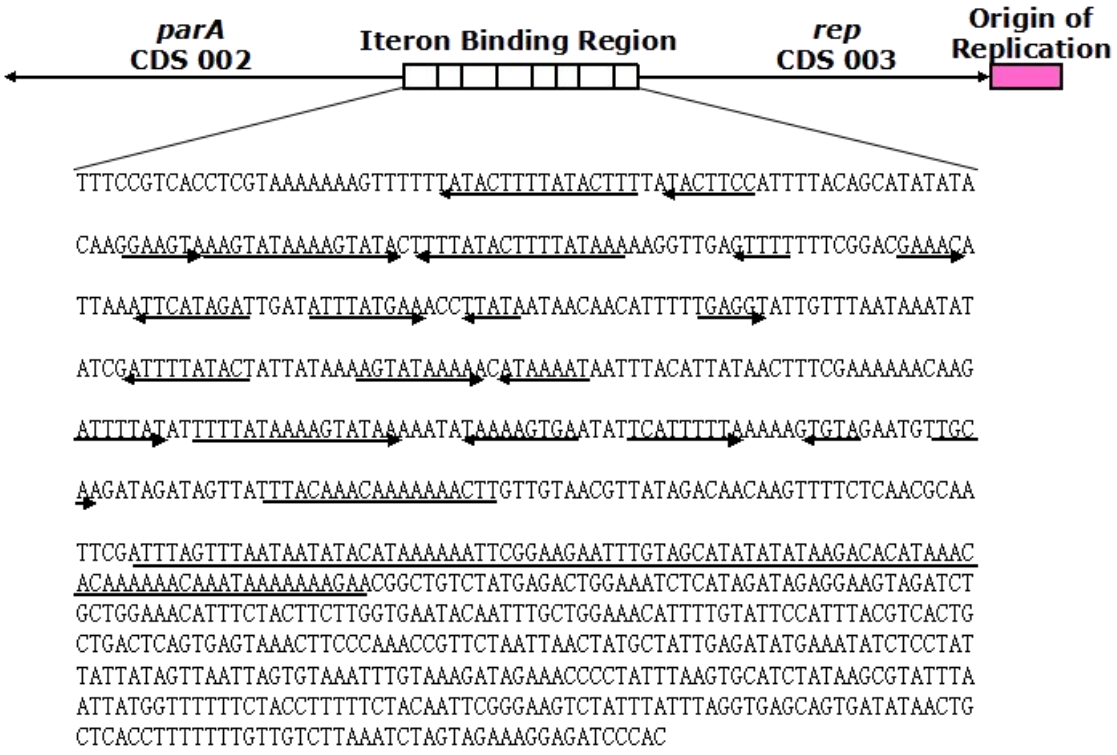


Fig. 20. The pMOGI222 replication region. The pMOGI222 iteron-binding region is between the copy number control gene *parA* and the replication gene *rep*. Arrows on the nucleotide sequence identify the iteron-binding repeat. The conserved regions of high A+T content are underlined.

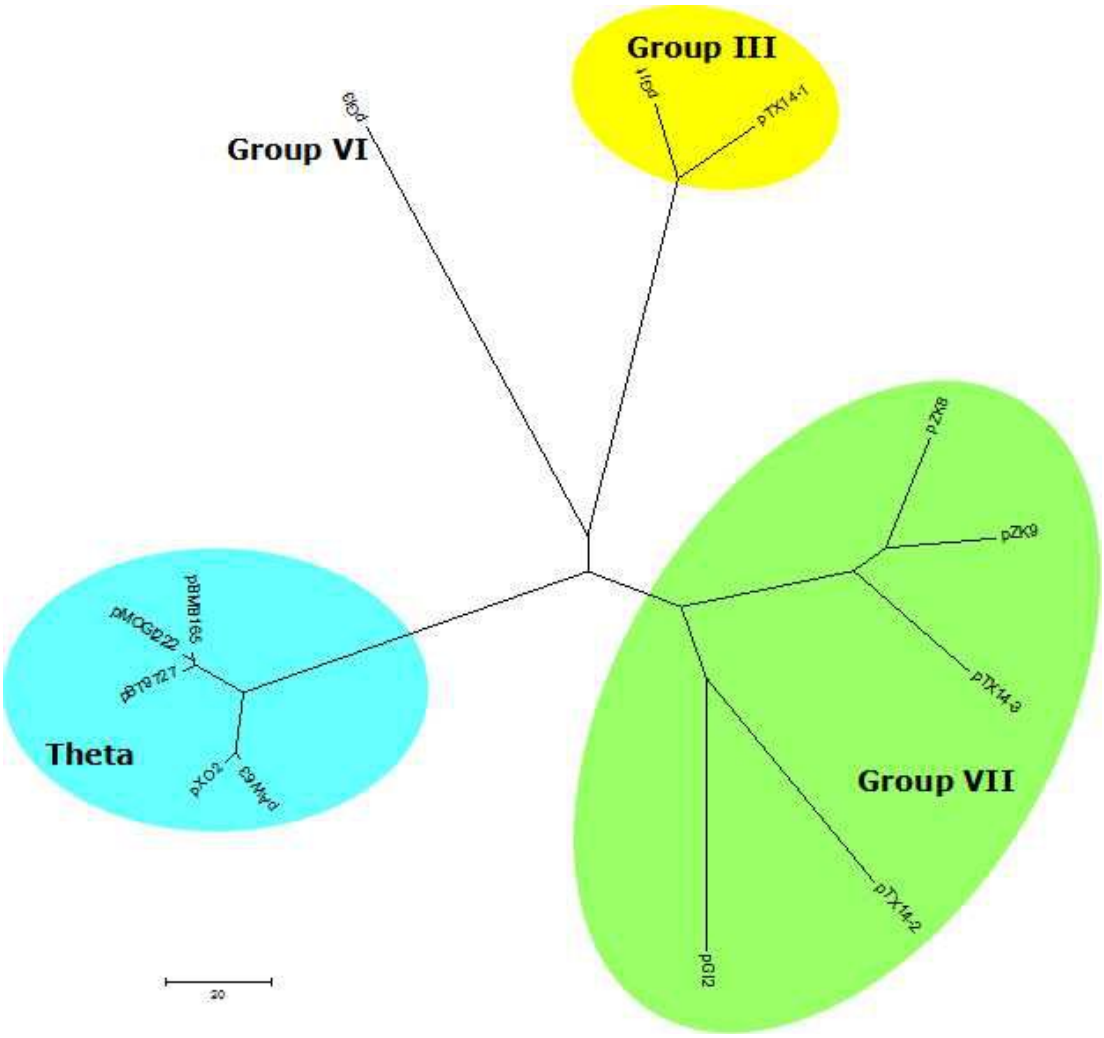


Fig. 21. Comparison and clustering of *B. cereus* group plasmids based on replication. Replication, as identified by annotation, were aligned and compared using CLUSTALW. The unrooted neighbor-joining phylogenetic trees were generated and displayed with MEGA 5.2. Theta-replicating plasmid proteins are within the blue ovals, Rolling circle replication plasmids of group VII are in the green ovals and only two members of the group III family replication proteins could be identified for inclusion.

4. Discussion

The completion of the genome sequence of *B. thuringiensis* subsp. *mogi* and comparative analysis with other bacteria species offers new information regarding the evolution of this species. Based on overall nucleotide and protein similarities, *B. thuringiensis* subsp. *mogi* is most similar to *B. cereus* G9842. The similarities including the synteny of genomes and a much similar megaplasmid, suggest that these two species are a biologically and phylogenetically divergent group whose members have developed to adapt to particular environmental conditions over evolutionary time.

The species in *B. cereus* group were classified as distinct because of the great relevance of their phenotypical differences, which formed the basis for their classification. *B. anthracis* strains are capable of capsule formation and the production of toxins that led to carbuncles in animals and in humans, causing the disease known as anthrax. *B. thuringiensis* forms a parasporal crystal that is active on larvae of a variety of insect orders. *B. cereus* lacks both of those characteristics and can cause food contamination. However, comparison of their 16S rRNA revealed less than 1% divergence between them (Ash *et al*., 1991).

In addition, many studies discovered that members of the *B. cereus* clade are very closely related in terms of phylogenic evolution as *B. thuringiensis* may resemble *B.*

cereus when losing its characteristic plasmids (Carlson *et al.*, 1994; Helgason *et al*., 1998; 2000), and vice versa, a *B. cereus* strain may display characteristic functional properties of *B. thuringiensis* or *B. anthracis* when it acquires plasmids of *B. thuringiensis* or *B. anthracis*, such as pBtoxis (Hu *et al.*, 2005) or pXO1/pXO2 (Hoffmaster *et al*., 2004).

On the other hand, despite the high identity of *B. thuringiensis mogi* and *B. cereus* G9842, there are two megaplasmids harboring several *cry* genes that are unique in *B. thuringiensis* subsp. *mogi.* The plasmid analysis of species belonging to the *B. cereus* group has become more important over the past few years due to its direct relationship with the pathogenic phenotype and the mobilization capabilities of these extrachromosomal elements (Hoffmaster *et al*., 2004; Rasko *et al*., 2005). It is crucial that the mechanisms implicated in replication, maintenance and gene transfer be understood, especially those involving toxin coding plasmids, which can affect public health. The variation of the chromosomes might be due to duplication and a complex and dynamic evolutionary process.

Chapter 3. Molecular cloning and characterization of mosquitocidal protein genes from *B. thuringiensis* subsp. *mogi*

ABSTRCT

To investigate the role of six three-domain *cry* genes (including *cry19Bb1, cry73Aa* with *cry40orf2, cry20Bb1, cry27Ab1, cry4Aa* and *cry56Ba1* with *cry39orf2*) in crystal production of *B. thuringiensis* subsp. *mogi*, the transcription level of these toxin genes were analyzed by quantitative PCR (qPCR). The results clearly indicated that all of these *cry* genes were successfully transcribed in wild type *B. thuringiensis* subsp. *mogi* strain in different expression time with different maximum levels. These *cry* genes were cloned to the *Escherichia coli-B. thuringiensis* shuttle vector, pHT1K, under the control of its own promoter, then introduced into an acrystalliferous *B. thuringiensis* Cry$^-$B strain for further molecular characterization. Another vector p1KSD, which containing a strong chimeric *cyt1Aa* promoter combined with the STAB-SD sequence was constructed and used to over-express the *cry* genes. To determine the function of the *cry39orf2* and over-express the *cry56Ba1* in *cry56Ba1* operon, different combinations of Cry56Ba1 and Cry39ORF2 were synthesized in strain Cry$^-$B. The stable inclusions in recombinant cells suggests that Cry39ORF2 assists in synthesis and crystallization of Cry56Ba1 by functioning like the C-terminal domain characteristic of Cry protein in the 130 kDa mass range. In addition, the

increased Cry56Ba1 yield under the *cyt1A-p*/STAB-SD promoter has broadened the possibility of application in other toxins.

Key words: *B. thuringiensis,* three-domain *cry* gene, quantitative PCR, over -expression.

1. INTRODUCTION

The mayor determinants of the insecticidal properties of *B. thuringiensis* bacteria are the δ-endotoxins produced during bacterial sporulation, which form two multigenic groups, *cry* and *cyt* (de Maagd *et al*., 2001). The nomenclature of Cry proteins is based on their primary sequence identity, and they have been classified in 70 subgroups. Different proteins not related phylogenetically form part of this classification, such as the group of three domain (3d-Cry) toxins, the mosquitocidal-like (Mtx-like) Cry toxins and binary-like (Bin-like) Cry toxins. The Mtx-like and the Bin-like toxins have similarity with the Mtx or Bin toxins produced by *B. Sphaericus*, although in the case of *B. Sphaericus* these toxins are toxic against mosquitoes whereas in *B. thuringiensis* they are toxic against coleopteran larvae (de Maagd *et al*., 2003).

Some similarity have been found in the 3d-Cry proteins. Domain I (a 7 α-helical bundle) is equipped for pore formation in insect epithelial membrane. Domain II (a triple β-sheet structure) may be responsible for receptor recognition. Domain III (a

β-sandwich region) may protect the toxin from further degradation during proteolytic processing or moderate toxin bilayer and toxin-toxin interactions (Honee and Visser, 1993; Vontersch *et al.*, 1994).

Among these toxins, the lineage of 3d-Cry toxins represents the largest group with more than 53 different subgroups of Cry toxins (Crickmore *et al*., 2011). One particular feature of the members of the 3d-Cry group is the presence of protoxins with two different lengths, 65 and 130 kDa. The main difference between the 65 and 130 kDa 3d-Cry toxin is the C-terminal extension that is found in the 130 kDa protoxins and is dispensable for toxicity, as it is cleaved by proteases present in the larval midgut (de Maagd *et al.*, 2001).

Since the first cloning of the *cry1Aa* gene from *B. thuringiensis* subsp. *kurstaki* HD-1 (Schnepf and Whiteley, 1981), more than 100 *B. thuringiensis* toxin genes have been cloned until 1981. The great number of sequences known to date is mostly a result of the strong interest in finding novel Cry proteins, with the focus on three main purposes: (i) the search for a new range of activities, (ii) the search for higher levels of toxicity, and (iii) the search for alternative toxins in case of resistance development (Noguera and Ibarra, 2010).

According to the genome sequence of the *mogi* described in Chapter 2, there are at least 17 toxin-related genes existed in *B. thuringiensis* subsp. *mogi,* and only the *cry* genes which possessed three domains were chose to do the further research in this

chapter. The search for novel Cry toxins has followed this strategy, construct a recombinant vector and expression it in acrystalliferous mutants of *B. thuringiensis* strain. Meanwhile, the conspicuously increase in Cry protein yield obtained with *cyt1A* promoters combined with STAB-SD sequence suggested that this combination might be useful for increasing yield of other toxins.

2. MATERIALS AND METHODS

2.1 Bacterial strains and growth media

The novel serogroup *B. thuringiensis* subsp. *mogi* (H3a3b3d) strain used in this research, was isolated from fallen leaves, sampled in a forest region of the city of Mungyeong, as previously described (chapter I) (Roh *et al*., 2009). Cloned *cry* genes and recombinant plasmid constructs were amplified in *Escherichia coli* Top10 [F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu) 7697 galE15 galK16 rpsL(Str^R) endA1 λ^-]. Cry^-B, an acrystalliferous *B. thuringiensis* subsp. *kurstaki* strain, used as the host for expression of the *cry* genes, was kindly provided by Dr. M. Ohba (Institute of Biological, Faculty of Agriculture, Kyushu University, Japan).

B. thuringiensis was grown at 28ºC with vigorous shaking in SPY medium for DNA preparation and GYS medium for expression of crystal proteins (Kronstad *et al*., 1983; Li *et al*., 2002; Nickerson and Bulla, 1974). The LB medium was used as a

primary culture of *B. thuringiensis* and in *E.coli* culture for plasmid preparation.

2.2 Preparation of RNAs for reverse transcription PCR of target *cry* genes

B. thuringiensis strain was grown in LB. Total RNAs were isolated from 2 ml portions of wild type *B. thuringiensis mogi*. From each aliquot taken during the sporulation phase, 2 ml were processed by precipitating their cell content, resuspending the pellet and washing several times by centrifugation in TE buffer (Tris 50 mM, EDTA 10 mM, pH 8.0). Last pellet was resuspended in 500 μl TES buffer (Tris base 50 mM, EDTA 10 mM, sucrose 20%, pH 7.5) supplemented with 10 mg lysozyme, and incubated at 37°C and 75 rpm for 30 min. Suspension was then pelleted at 4000 rpm for 10 min and treated with 1 ml Trizol (Invitrogen) to extract RNA according to the manufacturer's instructions. Total RNA (~1 μg) was then treated with 2 U DNAse I (Rnase-free, Takara) and evaluated by PCR amplification. Once negative PCR amplification was corroborated, specific fragments were amplified by RT-PCR when using bioneer RT-PCR premix, 150 ng total RNA, 100 ng primers (see Table 9) reaction mixture. Amplification conditions were: an initial step at 42°C for 60 min, and denaturation of 5 min at 94°C, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 57°C for 30 s, and polymerization at 72°C for 1 min, each cycle. Amplification finished with an extension step at 72°C for 5 min. Amplification products were visualized in 0.8% agarose gels.

Table 9. Nucleotide sequences of forward and reverse primers used for amplification of the target *cry* genes in reverse transcription- polymerase chain reaction (RT-PCR).

Primer	Sequence (5'-3')	Target gene
Fw-19RT	CGCTGCATGGAAACAAAATA	*cry19Bb1*
Re-19RT	CCAAAAAGGGACCATTCTCA	
Fw-73RT	CGTGATGCTAGTATGTTTGG	*cry73Aa*
Re-73RT	CATCTCGGTTCTATCTAGCC	
Fw-20RT	CCTCAAAATGGGCAACAAGT	*cry20Bb1*
Re-20RT	CGAATAGGATCATACGTTGG	
Fw-27RT	TTGGAAGAAAAACCCGAATG	*cry27Ab1*
Re-27RT	TTCCATTTGTTCTAACCGCC	
Fw-4RT	CTCAACCTTATGCAGATTGTGAT	*cry4Aa*
Re-4RT	TGCGGCTTGTGCATAAGT	
Fw-56RT	CCGGGTTATGAAGCCTTAC	*cry56Ba1*
Re-56RT	GTCCTTCAAGTGGCTCTCC	

2.3 Analysis of mRNA expression by qPCR

For monitoring the expression levels of different *cry* genes in *B. thuringiensis mogi*, qPCRs were conducted as follow. Total RNAs (1 μg) were reverse transcribed to cDNA with the QuantiTect® Reverse Transcription Kit (Qiagen) according to the manufacturer's instructions. For qPCRs, primers specific to *cry19Bb1, cry73Aa, cry20Bb1, cry27Ab1, cry4Aa, cry56Ba1,* and 16S rRNA were designed (Table 10). The 16S rRNA qPCRs were performed as internal controls to normalize the amount of RNA input. qPCR was carried out using the Bio-Rad CFX Manager real-time detection system (Bio-Rad). Reaction mixtures (20 μl) contained 10 μl 2×EvaGreen qPCR mastermix (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 0.2 mM each dNTP, 6 mM $MgCl_2$), 4 pM each primer and 25 ng of diluted cDNA. The thermal cycling conditions were denaturation at 95°C for 5 min, followed by 40 cycles of 95°C for 30 s, 58°C for 30 s, and 72°C for 30 s. This was followed by a melting curve program of 55 to 95°C with a heating rate of 0.5°C and final cooling at 10°C. The data obtained from qPCR were analyzed by relative quantification using the ΔC_T (cyclic threshold) method. One-way analysis of variance (ANOVA) was used to compare the mean differences of relative quantification in different strains at each time point with a significance level of $P < 0.05$. Samples were collected from three independent experiments.

Table 10. Nucleotide sequences of primers used for amplification of the target *cry* genes in qPCR.

Primer	Sequence (5'-3')	Target gene
Fw-16SrRNA	CGTAGGTGGCAAGCGTTATCC	*16S rRNA*
Re-16SrRNA	TCCTCTTCTGCACTCAAGTCTCC	
Fw-19Q [a]	ATCCGTTTACATCTATCCCCATCTCAC	*cry19Bb1*
Re-19Q	CCGCTCAAGCATACCACATTCG	
Fw-73Q	CGCCGAGTCGCATACTATGATTTC	*cry73Aa*
Re-73Q	GTCTCCCTGTCCTACGCCAAAG	
Fw-20Q	ACAGTCTGGATCTACCACACCTTG	*cry20Bb1*
Re-20Q	AACTCCGCCATAACTTCGTTGTATAG	
Fw-27Q	TGGTAACACGCTACGGTAAGGAG	*cry27Ab1*
Re-27Q	GGTTCATCACAGACTGGGACAATG	
Fw-4Q	CTCAACCTTATGCAGATTGTGAT	*cry4AA*
Re-4Q	GCTAGAACTGGCGCTGCTAT	
Fw-56Q	GCTCCAATTACTGGCGGAACATC	*cry56Ba1*
Re-56Q	ACTCTCAGCATTAGCGGGAAAGG	

[a] The superscript letters Q indicate primers for quantitative real-time PCR.

2.4 PCR amplification *cry* genes, cloning and sequence analysis

All PCR amplifications were performed with the DNA Thermal Cycler (Perkin Elmer Co., USA). For amplification of the active region of *B. thuringiensis*, the PCR was performed using an Expand Long Template PCR system (Roche Diagnostics GmbH, Germany) and a GeneAmp 2400 PCR system thermocycler (Perkin Elmer, Boston, MA) for 33 cycles. For PCR, 0.1 μg of total DNA from *B. thuringiensis mogi* was mixed with a solution containing each primer (Table 11) at a concentration of 20 pM, each dNTP at a concentration of 10 mM, and 1.25 unit of PrimeSTAR GXL DNA polymerase in PCR buffer (Takara, Japan). Conditions were as follows: 94°C for 1 min, 55°C for 1 min, 68°C for 4 min, and then 7 min of termination at 68°C. Then the PCR products were cloned into the pHT1K plasmid vector and sequenced (Fig. 22).

The expression vector (pHT1K, contained an erythromycin resistance gene marker) used in this study was designed to express the different *cry* genes in *Bacillus* spp. This vector encoded the insecticidal *cry* genes of *B. thuringiensis* subsp. *mogi* under the control of its endogenous promoter in a minimal *E. coli-B. thuringiensis* shuttle vector. All of these recombinant plasmids were transformed into *E. coli* Top10 and confirmed by nucleotide sequencing.

2.5 Transformation of *B. thuringiensis*

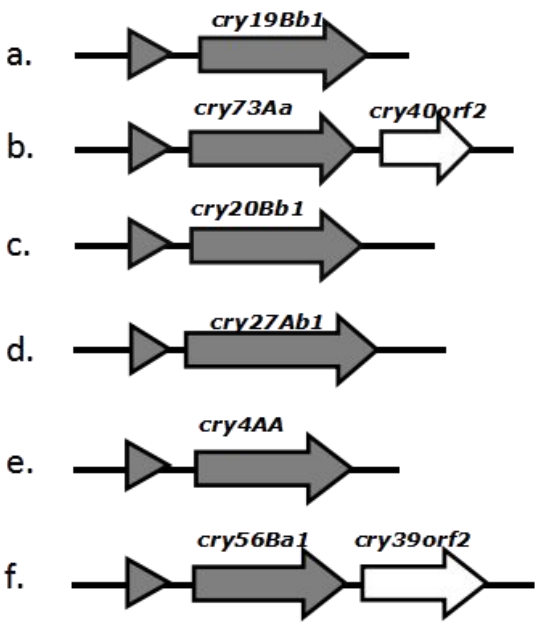
Once confirmed the sequences, these new constructs were used to transform the acrystalliferous *B. thuringiensis* subsp. *kurstaki* strain Cry$^-$B by electroporation following the protocol (25 μF, 2 kV, 400 Ω) commonly used for the transformation of *B. thuringiensis* (Lereclus *et al*, 1989) in 0.2 cm electrode gap electroporation cuvettes (Bio-Rad, USA). After the pulse, the electroporated mixture was added to 1 ml of brain heart infusion (BHI) (Beckton, Dickinson and Company, Sparks, MD) and incubated with gentle shaking (60 rpm) for 1 h at 30°C. Transformants were selected on nutrient agar plates containing erythromycin (25 μg/ml).

Table 11. Nucleotide sequences of primers used for amplification of the target *cry* genes to pHT1K vector in PCR.

Primer [a]	**Sequence (5'-3')**	**Target gene**
Fw-19	ACGAATTCGAGCTCGCATATTACAGCTTTTCCAC	*19Bb1*
Re-19	CCAGTGCCAAGCTTGCAGTTCACTCTTTTCTG	
Fw-73	CCATGTGCAGTTGTTGTATC	*73Aa+40orf2*
Re-40	CAAGGTTCGTCGACAATTC	
Fw-20	GTATACTGAGCTCTTGATTTCG	*20Bb1*
Re-20	GAATAGAGGCCATGCATGC	
Fw-27	CAAATCATGGTACCCAGCTAC	*27Ab1*
Re-27	GACATAGAACGTCGACACTTGG	
Fw-4	ACGAATTCGAGCTCGGAATCGCGCATATTTCAGG	*4Aa*
Re-4	CCAGTGCCAAGCTTGGAGCCAACTAGATATG	
Fw-56	GATAGAGCTCCATTGGTTATTGAAC	*56Ba+39orf2*
Re-39	CGTGTCGACAGATACCTCGTCAG	

[a] Primers for cloning target genes to pHT1K vector.

A



B

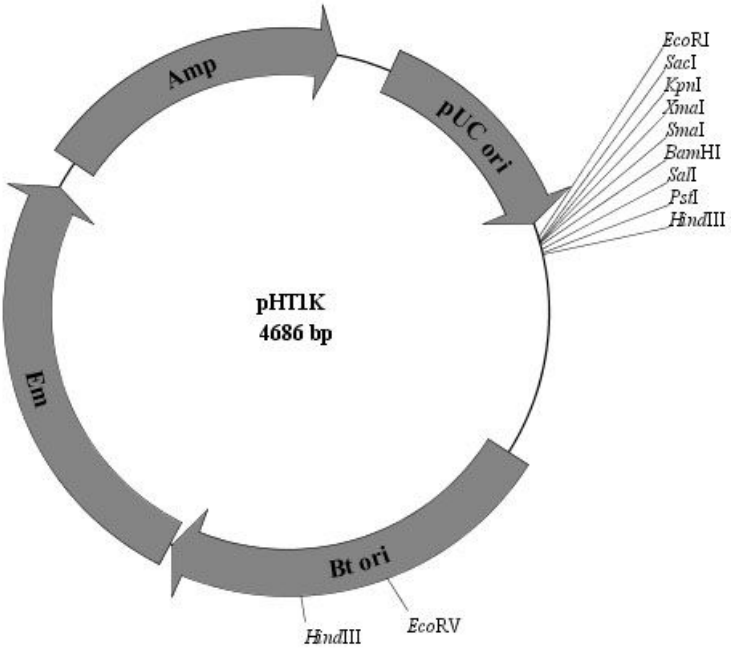


Fig. 22. Cloning of mosquitocidal *cry* genes from *B. thuringiensis* subsp. *mogi*. (A) The arrangement of six *cry* genes from *mogi.* (B) The *E. coli-B. thuringiensis* shuttle vector, pHT1K. Amp, amplicillin resistant gene; Em, erythromycin-resistant gene;

pUCori, *E. coli* Replication origin, Bt ori, *B. thuringiensis* replication origin.

2.6 Construction of recombinant plasmids for over expression of Cry56Ba1 and functional analysis of Cry39ORF2

The primers used to amplify different regions of the *cry56Ba1* operon are listed in Table 12, and the plasmids constructed using the PCR products are illustrated in Fig. 23. A series of recombinant plasmids were constructed for two different purposes:

(i) Analysis of the *cry56Ba1* operon

To determine the function of Cry39ORF2 encoded by the second gene in the *cry56Ba1* operon (Fig. 22A, panel f), DNA fragments containing *cry56Ba1* and *cry39orf2* were obtained using PCR, with *mogi* DNA as the template. The primer sets Fw-56 + Re-39 and Fw-56 + Re-56 (Table 12) were used to amplification fragments *cry56Ba1* + *cry39orf2* (Fig. 23, panel 1) and *cry56Ba1* only (Fig. 23, panel 2). Both of these products were cloned to pHT1K vector (Fig. 22B) and expressed under their original promoters.

(ii) Improvement of Cry56Ba1 synthesis, crystal topology, and determination of Cry39ORF2 function

To improve the synthesis of Cry56Ba1, plasmids containing the following elements were constructed as follows: fragment *cry56Ba1*+*cry39orf2* (Fig. 23, panel 3); *cry56Ba1* (Fig. 23, panel 4) and *cry39orf2* only (Fig. 23, panel 5). The primer sets which used for amplification the different constructs were Fw-56inf + Re-39inf, Fw-56inf + Re-56inf and Fw-39inf + Re-39inf, respectively (Table 12). All of these

products were cloned to p1KSD (Fig. 25) vector and expressed under the strong chimeric *cyt1A*-p/STAB-SD (Fig. 24) promoter.

All of these recombinant plasmids were transformed into *E. coli* Top10 and the integrity of these sequences was confirmed by restriction enzyme digestion and sequencing analysis. These new construct were then electroporated into *B. thuringiensis* Cry^-B as described before.

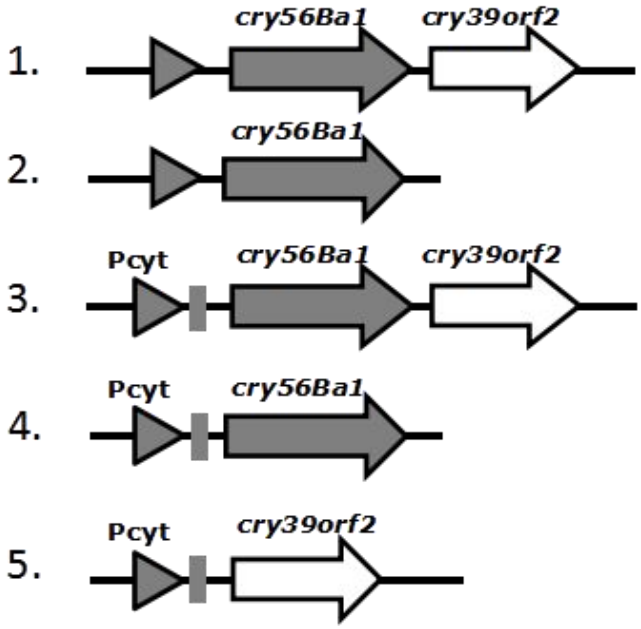


Fig. 23. Schematic illustration of the different constructs containing *cry56Ba1* or *cry39orf2* or both. Construction of No. 1 and 2, expression were under the control of the original wild-type promoter of *cry56Ba1*. Construction of No. 3, 4 and 5, expression were under the control of the c*yt1A*-p/STAB-SD sequence.

Table 12. Primers used for amplification of constructs of *cry56Ba1* and *cry39orf2*.

Primer	Sequence (5'-3')
Fw-56[a]	GATAGAGCTCCATTGGTTATTGAAC
Re-56[a]	CCAAGCTTGCATGCATAC
Re-39[a]	CGTGTCGACAGATACCTCGTCAG
Fw-56inf[b]	GAAGGCTTTTTCTAGTAAGGACTACATAAGGAGTG
Fw-39inf[b]	GAAGGCTTTTTCTAGTAGGTAGTAATCCTGTTC
Re-56inf[b]	GGCCAAGCTTGGATCTACCTACTTCATTACATAC
Re-39inf[b]	GGCCAAGCTTGGATCGCAATGATTTTAATTC

[a] Primers for cloning target genes to pHT1K vector

[b] Primers for cloning target genes to p1KSD vector.

Fig. 24. Upstream region and schematic illustration of constructs used to synthesize Cry proteins. A: Nucleotide sequence of the 660 bp fragment containing *cyt1A* promoters combined with STAB-SD sequence. Sigma E-like and sigma K-like promoters of *cyt1A* are shown, respectively, in boxes or underlined. The STAB-SD sequence is highlighted as a black box. B: Nucleotide sequence of the terminal sequence. Arrows indicate orientation of inverted repeat sites.

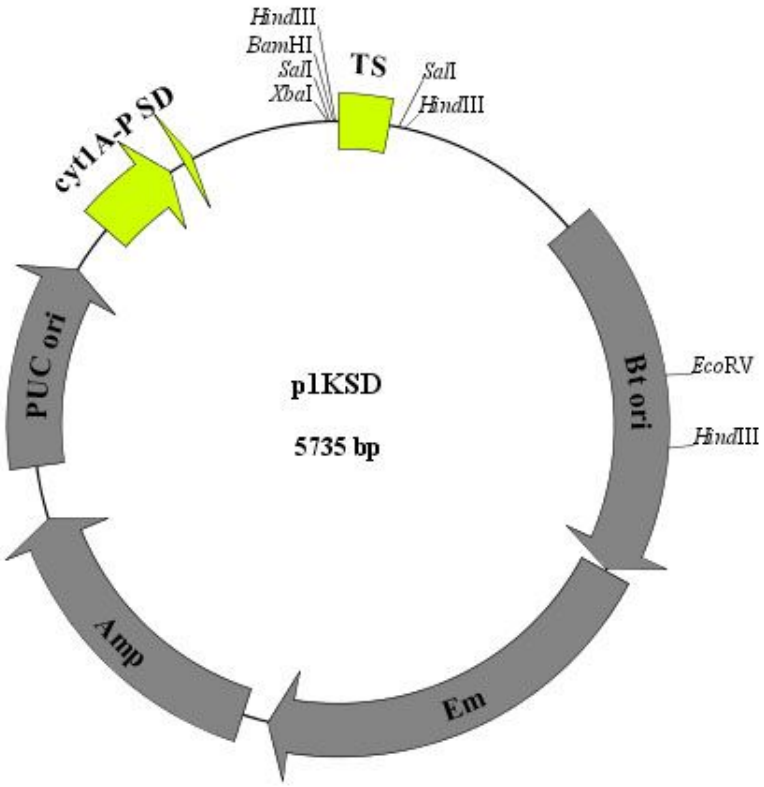


Fig. 25. The over-expression vector p1KSD. Amp, amplicillin resistant gene; Em, erythromycin-resistant gene; pUCori, *E. coli* Replication origin, Bt *ori*, *B. thuringiensis* replication origin.

2.7 Morphological observation and SDS-PAGE

Parasporal inclusions were purified by the method of Thomas and Ellar (1983). Crystal morphology of the isolates were examined by phase-contrast microscopy or transmission electron microscopy as described in Chapter 1 (see materials and methods 2.5). For SDS-PAGE samples, cells were cultured on NAE (nutrient agar plates containing erythromycin 25 μg/ml) medium plate at 28ºC and harvested after autolysis. SDS-PAGE was performed on a 12% separating gel with 5% stacking gel. The gel was stained with 0.1% Coomassie brilliant blue (Sigma Co., St Louis, MO, USA).

2.8 N-terminal sequencing

The N-terminal sequence of Cry20Bb1 and Cry56Ba1 proteins from recombinant *B. thuringiensis* Cry$^-$B strains were determined as follows. The inclusions from recombinant Cry$^-$B strains (CB/pHT1K-20Bb1 and CB/p1KSD-56Ba1+39orf2) were harvested and washed 3 times with a washing solution (0.5 M NaCl, 2 mM EDTA), then separated on a 10% of SDS-PAGE gel. The gel was transferred to a PVDF membrane (Amershem Hybond^{TM}-P, GE Healthcare). The membrane was stained with Coomassie brilliant blue R250 and the protein bands were excised and subjected to procise 491 HT protein sequencer (Applied Biosystems, USA) sequencing at the University of Korea, Korea.

2.9 Insects and toxicity assays

The mosquito larvicidal activities were assayed on 4th instar larvae of *Culex pipiens molestus* and *Culex pipiens pallens* (Diptera: Culicidae) which were grown in a container (35×25×3 cm) at 25ºC. Freeze-dried *B. thuringiensis* spores-crystal complex were suspended in double-distilled water. Suspensions were diluted to 6 or 7 different concentrations in cups in a final volume of 100 ml. Bioassays were replicated three times using 30 4th *Culex pipiens* instars of per concentration. After 48 h of exposure at 25°C, dead larvae were counted. Statistical analysis of data and 50% lethal concentrations (LC_{50}) were performed with probit analysis (Russell *et al*., 1977).

2.10 Amino acid sequence alignments

Homology searches with *cry* genes and orf2 were performed using the Basic Local Alignment Search Tool (BLAST) (BLASTP version 2.2.2; http: //www.ncbi.nlm.nih.gov/). ORF2 and Cry protein sequences that showed the highest level of identity were then aligned using MegAlign software (DNAstar, http://www.dnastar.com).

2.11 Nucleotide sequence accession number.

The nucleotide sequence data shown below (Fig. 26 - Fig. 31) are available in the GenBank, and the accession number listed in table 14.

3. RESULTS

3.1 Summary of toxin-related genes in *B. thuringiensis* subsp. *mogi*

According the CDS information from pMOGI364 and pMOGI222 (Chapter 2, table S1 and S2), there are at least 17 toxin-related genes existed in *B. thuringiensis* subsp. *mogi* (summary in Table 13). These two megaplasmids carry several insertion sequences and encode two further proteins (for example: CDS pMOGI364_340, which encodes a *cryBP1* family protein and shows homologue with P19; CDS pMOGI222_132, another 19 kDa accessory protein in *B. thuringiensis* subsp. *mogi*) with roles in promoting crystal formation and enhancing cell viability, probably by acting as chaperones (Dervyn *et al.*, 1995; Manasherob *et al*., 2001; Wu and Federici, 1993).

Six *cry* genes which contained intact 5 conserved blocks (including *cry19Bb1, cry73Aa* with *cry40orf2, cry20Bb1, cry27Ab1, cry4Aa* and *cry56Ba1* with *cry39orf2*) were chose for qPCR test and further studies. The detailed information about the nucleotide sequences and deduced amino acid sequence were shown in Fig. 26 to Fig. 31. These sequences were submitted to GenBank, and the accession numbers are given in Table 14.

As shown in Fig. 27, sequence of *cry73Aa* operon include two open reading frames oriented in the same direction and separated by 64 bp were identified (Fig. 27): *cry73Aa* (2,010 bp) and *cry40orf2* (1,494 bp). A putative ribosome binding site (RBS),

GGAGT, was identified 7 nucleotides upstream from the start codon of the *cry73Aa* gene (Fig. 27), and a sequence, AAAGGTTGTG, that could act as a RBS was identified 6 nucleotides upstream from *cry40orf2* (Fig. 27). The similar gene structure also emerge in *cry56Ba1* operon (Fig. 31), two open reading frames oriented in the same direction and separated by a 68 untranslated bp, two putative RBS sited were located 7 and 6 bp upstream of the star codon of *cry56Ba1* and *cry39orf2*, respectively.

Table 13. Summary of toxin-related genes in pMOGI364 and pMOGI222.

Locus_tag	Size (aa)	Strand	Annotation
pMOGI364_238	684	+	pesticidal crystal protein cry19Aa
pMOGI364_261	497	-	Cry8Ka2 delta-endotoxin
pMOGI364_262	669	-	Cry1-like delta-endotoxin
pMOGI364_287	340	-	mosquitocidal toxin gene
pMOGI364_322	299	-	35.8-kilodalton mosquitocidal toxin
pMOGI364_328	256	+	mosquitocidal toxin gene
pMOGI364_340	190	+	cryBP1 family protein
pMOGI364_348	722	+	Cry20-like delta endotoxin
pMOGI364_356	194	-	Cry20-like delta endotoxin
pMOGI222_106	220	-	mosquitocidal toxin
pMOGI222_108	825	+	Pesticidal crystal protein cry27Aa
pMOGI222_132	185	+	19kda accessory protein
pMOGI222_133	683	+	pesticidial crystal protein cry4AA
pMOGI222_134	506	-	crystal protein ET69
pMOGI222_186	557	-	mosquitocidal toxin protein
pMOGI222_235	562	-	Cry39ORF2 protein
pMOGI222_236	659	-	Cry56Aa-like protein

Table 14. Amino acid sequences of putative crystal proteins of *B. thuringiensis* subsp. *mogi* strain.

Protein No.	Locus_tag[a]	Protein Name	Predicted size	Protein size	GenBank accession No.
1	pMOGI364_238	Cry19Bb1	683 aa	78 kDa	KC182376
2	pMOGI364_262	Cry73Aa	669 aa	77 kDa	KC182375
	pMOGI364_261	Cry40ORF2	497 aa	57 kDa	KC182377
3	pMOGI364_348	Cry20Bb1	753 aa	83 kDa	KC182372
4	pMOGI222_108	Cry27Ab1	825 aa	94 kDa	KC182373
5	pMOGI222_133	Cry4Aa	683 aa	77 kDa	
6	pMOGI222_236	Cry56Ba1	659 aa	73 kDa	KC182374
	pMOGI222_235	Cry39ORF2	562 aa	64 kDa	KC182378

[a] locus tag means the *cry* genes located site.

Translation map – *cry19Bb1*

```
                                                                  -35
CATATTACAGCTTTTCCACTCCACAATCCACATGTGGATTACGATTATCCAGCTACTATA      60
                  -10
TCTATAAAATTTTGTATAATGAGATTAATTAAAAGATGTGTAAAGGCTGAATAATCCCCT     120
GGATACAAACATACGTATAAGATTGAAGAATAGTATCGATTCTGTCAATTACTTAGATAG     180
                          RBS                M  H  P  Y  Q  N  K  N  E
ATACTTCTTTAATATGAGCACGGGGGGAACTAAGATGCATCCTTATCAAAATAAGAATGA     240
Y  E  I  L  D  A  T  Q  N  N  C  H  M  S  N  C  Y  P  K  Y
ATATGAAATCTTAGATGCTACACAAAATAATTGTCACATGTCTAATTGTTATCCCAAGTA     300
P  L  A  N  D  P  Q  M  Y  L  R  N  T  H  Y  K  D  W  I  N
CCCACTAGCAAATGATCCTCAAATGTATTTGCGCAACACCCATTATAAGGATTGGATAAA     360
M  C  E  E  A  S  Y  A  S  S  G  P  S  Q  L  L  K  V  G  G
TATGTGCGAGGAAGCTTCTTATGCATCTTCAGGTCCTTCACAATTACTTAAAGTTGGAGG     420
S  I  V  A  K  I  L  G  M  I  P  E  V  G  P  L  L  S  W  M
TTCTATAGTTGCTAAAATTCTTGGAATGATTCCTGAAGTTGGTCCTCTTTTAAGTTGGAT     480
V  S  L  F  W  P  T  I  Q  E  K  N  T  V  W  E  D  M  I  K
GGTATCTTTATTTTGGCCAACTATTCAAGAAAAAAATACTGTTTGGGAAGATATGATAAA     540
Y  V  A  N  L  L  K  Q  E  L  T  N  Y  T  L  N  R  A  T  S
GTATGTAGCAAATCTGTTAAAACAAGAATTAACAAATTATACACTTAACCGTGCCACAAG     600
N  L  F  G  L  N  E  S  L  N  I  Y  N  R  A  L  A  A  W  K
TAATTTATTTGGATTAAATGAATCTTTGAACATATATAACCGAGCCCTCGCTGCATGGAA     660
Q  N  K  N  N  F  A  S  G  E  L  V  R  A  Y  I  N  D  L  H
ACAAAATAAAAACAATTTCGCAAGTGGGGAACTTGTAAGGGCATATATAAATGATCTTCA     720
I  L  F  T  R  D  I  Q  S  D  F  S  L  G  G  Y  E  A  V  L
TATACTCTTTACAAGAGATATTCAATCAGATTTCTCATTAGGAGGCTATGAAGCCGTATT     780
L  P  S  Y  A  S  A  A  N  L  H  L  L  L  L  R  D  V  A  I
ATTACCTTCATATGCAAGTGCTGCCAATCTTCATTTACTATTGTTACGTGATGTTGCAAT     840
Y  G  K  E  L  G  Y  P  L  E  D  V  E  F  Y  Y  N  E  Q  K
TTACGGAAAAGAATTAGGATATCCCTTAGAGGACGTAGAATTTTATTATAATGAGCAAAA     900
F  Y  T  E  K  Y  S  N  Y  C  V  N  T  Y  K  A  G  L  E  L
GTTCTATACAGAAAAATATAGTAATTATTGTGTAAATACGTACAAAGCGGGTTTAGAATT     960
A  K  Q  I  G  W  S  D  F  N  R  Y  R  R  E  M  T  L  S  A
AGCAAAACAAATAGGATGGTCAGATTTTAATCGTTATCGCAGAGAAATGACTTTATCCGC    1020
L  D  I  V  A  L  F  P  L  Y  D  T  R  L  Y  P  S  K  D  G
ATTAGATATAGTTGCCTTATTCCCACTGTATGATACAAGACTGTATCCGAGTAAAGATGG    1080
K  I  H  V  K  S  E  L  T  R  E  I  Y  S  D  V  I  N  A  H
TAAGATACATGTTAAATCTGAACTAACGAGAGAAATTTACTCTGATGTTATTAATGCTCA    1140
```

V D L V L K E D K A Y F T Q V E A L Y T
TGTAGACTTAGTCTTAAAAGAAGATAAGGCATATTTTACGCAAGTTGAAGCGCTTTATAC 1200
R R P H L F T W L R G F R F V T N S I S
ACGTCGACCACATTTATTTACTTGGTTACGAGGATTTAGATTTGTAACCAATTCCATTTC 1260
S W T F L S G A Q N K Y S Y T S S S S I
TTCTTGGACATTCTTATCTGGCGCGCAAAATAAATATTCTTATACATCCTCTAGTTCAAT 1320
E N G P F L G Q D T E Y G G T S S N M V
TGAGAATGGTCCCTTTTTGGGTCAGGATACAGAATATGGTGGAACCTCTTCTAATATGGT 1380
I P E N Q Y I Y N L W T K N Y E W I Y P
CATTCCAGAAAATCAATATATTTATAATTTATGGACCAAAAATTATGAATGGATTTACCC 1440
W T D P V N I T K I N F S L T D N N S S
TTGGACTGATCCAGTAAATATTACAAAAATTAATTTTTCTCTAACAGATAATAATTCCTC 1500
N E V I Y G A E R I N K P T V R T D F N
TAACGAAGTTATCTATGGTGCAGAAAGAATTAATAAACCTACTGTTCGAACAGATTTCAA 1560
F L L N K E G T G P A T Y Y D Y N H I L
TTTTCTGCTTAACAAAGAAGGCACTGGTCCTGCGACATATTATGATTATAATCACATTTT 1620
S Y T L I N G S T A G Q K R H G Y S F A
ATCATATACACTAATAAACGGAAGTACTGCTGGTCAAAAAAGGCACGGATATTCATTTGC 1680
F T H S S V D P Y N K I A T D K I T Q I
TTTTACaCATAGTAGTGTGGACCCATATAACAAAATTGCCACAGATAAAATTACTCAAAT 1740
P A V K S N G W M F F G D V L K G P G H
TCCTGCGGTGAAAAGTAATGGATGGATGTTTTTTGGTGACGTATTAAAAGGTCCTGGCCA 1800
T G G D L V T L S N G G R Y T L N I I F
TACAGGTGGAGATTTAGTGACTCTTAGTAATGGGGGTAGATATACACTAAATATTATTTT 1860
P A Q A Y H I R I R Y A S N G D G E M G
CCCCGCTCAAGCATACCACATTCGTATTCGGTATGCTTCTAATGGTGACGGTGAGATGGG 1920
I D V N G S G Y T R F S I K S T F S H N
GATAGATGTAAACGGATCGGGGTATACCCGTTTTAGTATAAAGAGCACTTTTTCTCATAA 1980
N Y N D L N F Q D F N L M D T S F I Y N
TAATTATAATGATTTAAACTTCCAAGATTTCAATTTAATGGACACATCTTTTATTTACAA 2040
A T Y T G S Q T I W L Y S Y A T A R V I
TGCAACTTATACAGGATCACAGACTATATGGTTATACAGTTATGCAACAGCACGGGTAAT 2100
I D K I E F I P V G T F A N Q L L E E T
TATAGATAAAATTGAATTTATACCAGTTGGTACTTTTGCAAATCAATTATTAGAAGAAAC 2160
Q C Y N Y N Q N M D N T Y Q P S Y A N T
ACAATGTTATAACTATAATCAAAACATGGATAATACATACCAACCAAGCTATGCCAATAC 2220
Y N H N S S N M H N Q S Y N N *
CTACAATCACAACTCAAGTAATATGCATAATCAAAGTTATAACAATAATTATAACCAAAA 2280
IR IR
CATGGATAATACATACCAACCAAGCTATGACAATACCTACAATCACAACTCAAGTAATAT 2340

GCATAATCAAAGTTATAACAATAATTATAACCAAAACATGGATAGCATGTACAACAATAA 2400
CTATAGCCAAAGTACTAATGATATGTACCCTCAAGAATATACTAACAGCAATAACCAAAA 2460
ATTCGGCTGTACATGTAATCAAGGGTATAATAACTATCCAAAATAAGTACAGAAAAGAGT 2520
GAACTGCACCCCAATTGTTAGACACAGTTTAACAATTGGAGGTGCAGTTTTTCTATGGCT 2580
AAATTTACAGCTGATGAAAAAATACAAATCGTTCTACGTTATTTGAACGGAAATGAAAGT 2640
TATCGAGAAATGGGTAGATCGCTCGGTATAAGTGACACAATCATTTTGAATTGTGTAAAA 2700
CAGTGCTCAACAATTGCAAATTCTTTTATGTAGAGGAACTCCGACCTCAAAACCCCAAAG 2760
AATTTTTCTATCACAGCGTTGTCGTAACATTTTCCTTTTCGAGACATACTCTGGACGATA 2820
GCTCTAGATTCAAGTGT 2837

Fig. 26. Nucleotide sequence and deduced amino acid sequence of the *cry19Bb1* gene. The gene is 2,052 bp in length and codes for a polypeptide of 683 amino acids. The potential -35 and -10 boxes and a putative ribosome-binding site (RBS) are overlined. The stop codon is marked with asterisks. Five conserved sequence blocks (blocks 1 to 5) are shadowed. Terminal inverted repeats (IR) are indicated below the arrow.

Translation map – *cry73Aa+cry40orf2*

```
CGGCAATCTCAATTAAATCCTTGGTAGAACGTGAAAGTCGATTCAGGTATGTAACTGCTA    60
CTGTATCTCCATTCCGCAGAAACTTTTACATTCGATCTAGCTCCAGACGATCTCGTTTGG    120
TTCCTGTCGGTCTAAATTTGGATCAGTGGTAGATACCAGTATATATCCAAACTTCATATG    180
                    -35                 -10
ATTCCTTCTTTTCAGTTTTATTGTATCAAAAACGATATAAAATTCGTATTGTAGCAAATT    240
AATTTTGTTACAGTACGATTGCTAGTTTACTTAATTTCAGCTCTTGATACCTATCCTTGT    300
CACAAAAACGGAAATTCTTTGTACATAGGTAAAATTCTCGTGATTCGGTTTTCTCTAGAA    360
AAAATGAACTGAATCCAATACATGTTCAGATTTATTTTCTACAATTTCTTTCAATGATGA    420
ATAATTAAAAATCTACAACAATCTATTCAATATTCCTCTCTTTTTTCGAGGGATAGGAAA    480
AAACATCCAAGTGTGAATTTTGTACATATAAAGGTGAATAACTCTTTCCACATTCTAAAA    540
AAACAACAAAGAAAAAATCGTTCTACAGAAATCTGAAGCTTTTAAAAAAATACATACAAT    600
ACATAAAGAGAAGGTTTAAAAAAATAGATGCCTCACAAAATATAATGGGTTTATTTGTAG    660
AAACATCGTTAGAGGAATACATTGGGATGCTGCGAATATATAGAAAGACATCTAGTATAT    720
                                   RBS       M  N  S  Y  Q
ATTCATTAGGTATCTTAATATAAGGATTACATAAGGAGTGAAAATATGAATTCATATCAA    780
N  K  N  K  N  K  Y  E  M  L  D  T  S  R  K  S  S  N  M  S
AATAAAAATAAAAATAAATATGAAATGTTAGATACTTCAAGAAAAAGCTCTAATATGTCT    840
T  C  Y  P  R  Y  P  I  A  K  N  P  Q  K  T  M  Q  N  T  N
ACTTGTTATCCTCGGTACCCAATAGCAAAAAATCCACAAAAAACCATGCAAAATACGAAT    900
Y  K  D  W  I  N  M  C  T  S  K  N  L  E  D  G  I  Y  S  T
TATAAAGACTGGATAAATATGTGTACATCAAAAAATCTTGAAGATGGCATTTACTCTACA    960
S  A  K  D  V  I  T  N  S  I  N  I  S  S  Y  I  I  S  M  L
AGTGCAAAAGATGTAATTACAAATTCTATTAATATTTCTAGTTATATAATAAGTATGTTA   1020
G  M  P  Y  L  S  S  I  V  A  I  W  G  V  L  F  N  A  L  W
GGCATGCCTTACTTATCAAGCATCGTAGCTATATGGGGGGTGCTTTTTAATGCATTATGG   1080
P  S  S  D  N  Q  W  E  P  Y  M  N  H  V  Q  G  L  I  R  R
CCTAGTTCAGATAATCAATGGGAACCTTATATGAATCATGTACAAGGTCTTATTAGGCGA   1140
E  L  Q  T  F  A  R  E  Q  A  L  R  Q  L  E  G  L  G  G  N
GAATTACAGACTTTTGCAAGAGAACAAGCACTTAGACAATTAGAGGGGTTAGGTGGAAAT   1200
L  D  L  Y  K  E  A  L  E  E  W  E  Q  D  R  D  N  Q  T  T
TTAGATTTATATAAAGAGGCATTGGAAGAATGGGAGCAAGACCGTGACAATCAAACAACT   1260
K  E  R  V  R  D  R  F  R  I  L  D  G  F  F  T  Q  Y  I  P
AAAGAAAGGGTAAGAGATCGATTCCGCATATTGGATGGTTTTTTCACGCAATACATTCCG   1320
V  F  R  I  Q  G  Y  E  V  Q  L  L  S  V  Y  T  K  V  A  N
GTTTTTAGAATCCAAGGGTATGAAGTTCAATTATTATCTGTCTATACAAAAGTTGCGAAT   1380
L  H  L  L  L  L  R  D  A  S  M  F  G  A  D  W  G  M  S  Q
CTCCATTTGCTTTTATTACGTGATGCTAGTATGTTTGGGGCCGATTGGGGAATGAGCCAA   1440
T  N  I  N  D  N  Y  N  R  Q  M  N  L  T  S  L  Y  T  N  H
```

ACTAATATTAATGATAATTATAATCGACAAATGAACTTGACTTCATTATACACAAATCAT 1500
C V D F Y N Q G L N E A K A L S N S N W
TGCGTTGATTTTTATAATCAGGGTCTAAATGAAGCTAAAGCATTATCAAATTCAAATTGG 1560
D I F N D Y R R E M T I T V L D I V A L
GATATTTTTAATGATTATCGTAGAGAAATGACTATAACAGTTTTAGATATAGTTGCTCTA 1620
F S S Y D Y R R Y P I T T K V E L T R E
TTTTCTTCCTATGATTACCGTCGCTATCCTATAACTACAAAAGTAGAACTTACTAGAGAG 1680
I Y T P A I A S Q T W S N H N H L S P N
ATATATACACCTGCAATTGCTTCTCAGACTTGGAGTAATCATAATCATCTGTCACCGAAT 1740
I N F E F Y E N N L V R P P A F F T W L
ATAAATTTCGAATTTTACGAAAATAACCTTGTGCGGCCTCCTGCTTTTTTTACTTGGCTA 1800
D R T E M F S R Y L S T V V S E A W G G
GATAGAACCGAGATGTTTTCTAGGTACTTAAGTACTGTGGTTTCTGAAGCTTGGGGAGGG 1860
H I N H F H H T G E L P L S S R S G F I
CATATAAATCATTTTCATCACACTGGAGAACTACCTTTATCTTCACGTAGTGGTTTCATT 1920
G S D Q R R V A Y Y D F N F V G N D V F
GGAAGTGATCAGCGCCGAGTCGCATACTATGATTTCAATTTTGTTGGTAATGATGTCTTT 1980
R I Y S R V M S N P V G N Y F G V G Q G
CGTATATATTCAAGAGTAATGTCAAATCCAGTTGGGAACTACTTTGGCGTAGGACAGGGA 2040
D F Y L V N R D N C N T K T I T F T T K
GACTTTTATCTTGTGAATAGAGATAATTGTAATACTAAAACAATAACTTTTACTACTAAA 2100
A T N S N Q R S I L S E F P G E N S D P
GCAACGAATTCAAATCAAAGATCCATATTGTCTGAATTTCCGGGTGAAAATTCAGACCCA 2160
P T S K D Y S H R L S W I S G A F I G S
CCAACTTCTAAGGATTATAGCCATAGATTATCATGGATATCAGGTGCATTTATTGGTAGC 2220
D I A N V L S Y G W T H R S V D P N N T
GATATTGCAAATGTGCTTTCATATGGCTGGACTCATAGAAGTGTTGACCCTAACAATACT 2280
I Y P D K I T Q I P A V K L S S A S N C
ATTTATCCAGATAAGATTACTCAAATTCCGGCTGTTAAACTAAGTAGTGCTTCCAATTGC 2340
T V I P G P G S T G G H L V S F D R N G
ACTGTAATCCCAGGCCCTGGATCTACGGGAGGTCATTTAGTAAGTTTTGATAGGAACGGA 2400
S L D M Q F E F I T T Q T E Y R I R I R
AGTTTGGATATGCAATTTGAATTTATAACTACACAAACAGAGTATCGTATTCGTATACGC 2460
Y A S I A I N T L F F S F S G V N Q S I
TATGCCTCTATAGCAATAAATACACTATTCTTTTCTTTTAGTGGAGTAAATCAAAGTATA 2520
A L N S T G A S S L N N L R S E D F A Y
GCACTTAATTCTACAGGTGCTTCGTCACTAAATAATTTGCGAAGTGAGGATTTTGCGTAC 2580
L E F P Y G I F K P A I G N T L R I S N
TTGGAATTTCCGTATGGTATTTTTAAACCTGCTATAGGTAATACATTAAGGATTTCAAAC 2640
W S T V A P H L V I D K I E F I P I N S

TGGAGTACTGTTGCTCCACACTTGGTAATAGATAAAATTGAATTCATCCCAATTAATTCC 2700
T T A K Y E R M K E I E K A T T V V N S
ACTACTGCAAAATATGAGAGAATGAAAGAGATAGAAAAAGCCACAACAGTAGTGAATAGT 2760
L F I N *
TTATTTATAAATTAATACATGATATTAAAATACATGCCCTTGCTATTTAAAAATTGAACG 2820
RBS M F T N N A E N T L K I E T
GAAAAGGTTGTGGGAAATATGTTCACTAATAATGCGGAAAATACATTGAAAATAGAAACA 2880
T D Y E I D Q A A I S I E Y M S D E Q Y
ACAGATTATGAAATAGATCAAGCGGCTATTTCTATAGAATACATGTCGGACGAGCAATAT 2940
P Q E K M M L W E E I K H A K Q L S E S
CCACAAGAAAAAATGATGTTATGGGAAGAAATAAAGCATGCAAAACAACTTAGCGAATCT 3000
R N L L Q N G D F Q D S Y G Y G E N G W
CGTAATTTACTACAAAATGGAGATTTTCAAGATTCTTATGGGTACGGGGAAAATGGGTGG 3060
T N S N G I T I Q S N D P I F K G H Y L
ACAAACAGTAATGGTATTACCATTCAATCTAATGATCCTATTTTTAAAGGACATTATCTT 3120
Q M F G A R N I D G T L F P T Y I Y Q K
CAAATGTTTGGGGCAAGAAATATTGATGGAACGCTATTTCCAACCTATATCTATCAAAAA 3180
I D E F K L K P Y T R Y R V R G F V R S
ATAGATGAATTTAAATTAAAACCATATACACGTTATCGAGTAAGAGGATTTGTAAGAAGT 3240
S K D L K L V V T R Y G K E I D V I M D
AGTAAAGATTTAAAATTAGTGGTAACACGCTACGGTAAGGAGATTGATGTTATTATGGAT 3300
V P N D V A Y M Q P R H S C G D Y N R W
GTTCCAAATGATGTGGCATATATGCAACCACGTCATTCATGTGGAGATTATAATCGTTGG 3360
E S L S Q S V M N Q E Y P T P Y A A D A
GAATCATTGTCCCAGTCTGTGATGAACCAAGAATATCCTACACCATATGCAGCAGATGCC 3420
F D M Y S S Q F N R G K K H V T C H D C
TTCGATATGTATTCATCCCAGTTCAATCGAGGTAAGAAACATGTTACGTGTCACGATTGT 3480
H S F D F H I D I G E L D T N T N L G I
CATTCATTTGATTTTCATATTGACATAGGAGAATTAGATACAAATACAAACTTAGGTATT 3540
W V L F K I S N P D G Y A T L G N L E V
TGGGTCTTATTTAAAATTTCTAATCCAGATGGATACGCTACATTAGGCAATCTAGAAGTA 3600
I E E G P L T D E T L A H V K Q K E K K
ATTGAAGAAGGACCACTAACAGACGAAACATTAGCACATGTGAAACAAAAGGAAAAGAAA 3660
W N Q Q M E K K R C E T Q Q A Y N R A K
TGGAATCAACAGATGGAGAAAAAACGTTGTGAAACACAACAAGCCTATAATCGAGCAAAA 3720
Q A V D R L F T S T Q G E E L Q Y H I T
CAGGCAGTAGATAGATTATTCACAAGTACACAAGGAGAAGAATTACAATATCATATTACT 3780
L D H I K K S D Q L V Q S I P Y V H Q D
TTAGATCATATTAAGAAATCCGATCAGTTGGTACAGTCGATTCCCTATGTACATCAGGAT 3840
W L S D V P G M N A D L Y T D L N G R I

```
TGGTTATCAGATGTTCCAGGTATGAACGCTGATTTATATACAGATTTAAATGGACGTATA    3900
T  Q  A  R  Y  L  Y  D  A  R  N  I  I  T  N  G  D  F  T  Q
ACGCAAGCACGTTATTTGTATGATGCACGAAACATTATAACAAATGGTGATTTTACACAG    3960
G  P  T  G  W  S  A  S  G  H  E  A  F  K  K  I  D  G  D  S
GGACCAACAGGATGGAGCGCATCAGGACACGAGGCGTTCAAAAAAATAGATGGAGATTCT    4020
V  L  V  L  S  S  W  S  T  G  V  S  Q  N  L  H  V  Q  H  H
GTATTAGTTCTATCAAGCTGGAGTACCGGGGTATCTCAAAATCTGCATGTGCAACATCAT    4080
H  G  Y  V  L  R  V  I  A  K  K  E  G  L  G  K  G  Y  V  T
CATGGGTATGTATTACGTGTGATTGCGAAAAAAGAAGGGTTAGGAAAAGGATATGTAACG    4140
M  M  D  C  N  E  N  Q  E  T  L  K  F  T  S  C  E  E  G  Y
ATGATGGATTGTAATGAAAATCAGGAAACGCTTAAATTCACTTCTTGTGAAGAAGGATAT    4200
I  I  K  S  V  E  V  F  P  E  S  D  C  I  R  I  E  I  G  E
ATAATAAAATCAGTAGAGGTATTCCCAGAAAGCGATTGTATAAGAATAGAAATTGGAGAA    4260
T  E  G  T  F  Y  I  Q  S  I  E  L  L  C  M  K  G  Y  T  G
ACCGAGGGTACGTTTTATATACAAAGTATCGAGTTGCTTTGTATGAAAGGTTATACTGGA    4320
N  C  N  *                IR                    IR
AATTGTAACTAAAATACGCGTGCTATGTATGAACAAATATATAGTAGCAACTACAACCAT    4380
AATACTAGCGATAGGTATAATCAAAATTATACCAACAATTACGACCAGCATTCCAGCTGT    4440
ACGTGTAATCAAGAATATAACCGTTAAGATTCAAAATGAGGATCAGCATATTGACGAAAA    4500
AAATAAAAACCTACTCACAAAATCTATTGCGTATCATAACATAAGCTTTACAAATAGGAG    4560
ACATATTCTAGAACTGGTCTCCTTAATTCTAAAATAAGGAGGTCCTTTTCGTTTTCACAA    4620
TATCGATTAATGAAAATACTCCTTTACAGAACGATTTAGGCTGATTAGATTTGAATGTTG    4680
TTGAATTG                                                        4688
```

Fig. 27 Nucleotide sequence and deduced amino acid sequence of the *cry73Aa* and *cry40orf2* gene. The *cry73Aa* gene is 2,010 bp in length and codes for a polypeptide of 669 amino acids. The *cry40orf2* gene is 1,494 bp in length and codes for a polypeptide of 497 amino acids. The potential -35 and -10 boxes and a putative ribosome-binding site (RBS) are overlined. The stop codon is marked with asterisks. Five conserved sequence blocks (blocks 1 to 5) are shadowed. Terminal inverted repeats (IR) are indicated below the arrow.

Translation map – *cry20Bb1*

```
CTTGATTTCGAGGGAAACAAGCTAAATTATTATTTGAGTAAGAAACGAGATGCGAAGGCT    60
GCCAAGCGTTTTCTACAAAAACCCTTAGCTTCTTTTCATGTCACAAAAACTCGTGTCATA   120
ACTGCAGATGGTGATAAAGCCTATCCCGTTGTGATATAGAAATTAATAATGAAAAAAGCA   180
        -35                     -10
TACTGGTGGATAAATATAAATAAAACCTAATATATAAAAATAAATATACGAAGAGGGATT   240
CTAAAAAATCAATACTTTTACCAAAAATAATAGCTTTATTTGTAGAAAGATTATTTCAGG   300

AATGCATAGGCGCATTACGAATATTTACAAAGAAACCAATTATATATTTGTTAAAGGTAT   360
                    RBS          M  N  S  Y  K  N  N  H  T  M
TAAAAAATGTCTACATAAGGAGTGAAAAAAAATGAATTCTTATAAAAATAACCATACAAT   420
 V  N  S  P  E  N  S  S  N  T  V  N  R  Y  P  Y  A  C  N  P
GGTTAATTCCCCCGAAAATTCTAGCAATACTGTAAATAGGTATCCTTACGCTTGTAATCC   480
 N  I  E  T  Q  N  M  N  Y  K  D  W  M  A  G  Y  E  E  I  A
AAATATTGAAACGCAAAATATGAATTATAAAGATTGGATGGCTGGATATGAAGAAATTGC   540
 P  S  S  L  S  L  I  L  S  S  I  G  I  L  N  Q  V  I  A  L
TCCATCTTCATTATCTTTAATTTTATCTTCAATAGGCATTCTTAATCAAGTAATTGCCTT   600
 T  G  V  L  G  K  T  P  E  I  I  N  I  V  Q  E  M  V  S  L
AACTGGCGTATTAGGTAAGACACCAGAAATTATTAACATAGTACAAGAAATGGTGAGTTT   660
 I  R  G  N  T  G  N  D  L  L  V  H  V  E  Q  L  I  Q  Q  T
AATTAGAGGGAACACAGGCAATGATTTATTAGTACATGTAGAACAACTTATTCAACAAAC   720
 L  A  T  Q  Y  R  S  A  A  T  G  A  I  Y  G  I  S  R  A  Y
ATTGGCAACACAGTATAGGAGCGCTGCAACTGGAGCCATATATGGTATATCTAGAGCATA   780
 N  N  Y  L  E  F  F  R  Q  W  E  R  N  R  T  P  Q  N  G  Q
CAATAATTATTTGGAGTTCTTTAGGCAATGGGAACGTAATAGAACTCCTCAAAATGGGCA   840
 Q  V  E  S  A  F  T  T  V  N  T  L  C  I  N  A  L  A  P  Q
ACAAGTCGAGAGTGCTTTTACTACTGTTAATACTTTATGTATTAATGCTTTAGCCCCTCA   900
 A  S  L  S  R  R  G  F  E  T  L  L  L  P  N  Y  A  L  A  A
GGCGTCACTTTCACGCAGAGGATTCGAAACTCTTTTATTACCCAACTATGCTCTAGCGGC   960
 N  F  H  L  L  L  L  R  D  A  V  L  Y  R  T  Q  W  L  A  N
AAATTTCCATTTGTTATTATTAAGAGATGCTGTTCTTTATAGAACCCAGTGGTTAGCTAA  1020
 S  I  S  T  T  N  V  N  I  Q  I  L  T  R  A  I  N  E  Y  R
TTCTATTTCAACAACAAATGTAAATATCCAGATATTAACAAGAGCCATAAATGAATATCG  1080
 N  H  C  N  Y  W  Y  N  N  G  L  N  R  F  T  R  T  S  F  N
TAATCATTGTAATTATTGGTATAATAACGGATTAAATAGATTTACACGCACCTCTTTTAA  1140
 D  W  V  R  F  N  A  Y  R  R  D  M  T  L  S  V  L  D  F  V
TGATTGGGTTCGATTCAATGCTTATCGTAGAGATATGACGTTGTCGGTATTAGATTTTGT  1200
 T  V  F  P  T  Y  D  P  I  R  Y  P  R  P  T  N  V  E  L  T
TACAGTATTTCCAACGTATGATCCTATTCGATATCCAAGACCAACAAATGTTGAGTTGAC  1260
```

R I V Y T D P I S P P R G F P R T N P P

TAGAATTGTTTATACCGATCCAATAAGTCCACCTAGAGGATTTCCTAGAACCAATCCTCC 1320

S F N Q M E N L I I S G G P S F L N Q L

TTCCTTTAATCAGATGGAAAATTTAATTATTTCGGGTGGTCCTAGTTTCTTGAATCAATT 1380

R I Y T T F Y H D P H H V N R D Y W A G

GAGAATATATACAACTTTTTATCATGATCCTCATCATGTAAATAGAGACTATTGGGCCGG 1440

N R N Y L S N G I S R Q S G S T T P W P

GAATCGGAATTATTTAAGCAATGGGATTTCTCGACAGTCTGGATCTACCACACCTTGGCC 1500

T N I P M Q N I D I F R V N L T T H D I

AACTAATATACCTATGCAAAATATTGATATTTTCAGAGTGAATCTAACTACCCATGATAT 1560

D S I Q R S Y G G V H R S D F I G V N T

TGACTCTATACAACGAAGTTATGGCGGAGTTCATAGATCTGATTTCATTGGTGTAAATAC 1620

I N N Q R T T L F Y H Q N V D T S R F L

AATAAATAATCAAAGAACAACATTGTTCTATCACCAAAATGTGGATACTTCTCGTTTTCT 1680

T R N E T V F L P G D S G L E P N E Q N

AACAAGGAATGAAACAGTATTTTTACCAGGGGATTCAGGATTAGAACCAAATGAACAGAA 1740

Y T H R L F Q V M T T Y R I N P N A R R

TTATACTCACAGGTTATTTCAAGTCATGACCACATATCGTATTAACCCGAATGCTCGTAG 1800

A A F L H A W T H R S L R R R N G F R T

GGCAGCTTTTTTACATGCATGGACGCATAGAAGTTTAAGACGTAGAAATGGATTTAGGAC 1860

D Q I M Q I P A V K T I S T G D D R A V

GGATCAGATTATGCAAATACCAGCTGTTAAGACCATAAGCACTGGTGATGATCGTGCTGT 1920

V L N Y G E N I M K L D N L T S G L S Y

AGTGTTAAATTATGGAGAAAACATCATGAAATTAGATAATTTAACTTCAGGTTTATCCTA 1980

K V T A T D S A A S N T R F I V R V R Y

TAAAGTAACGGCGACAGATTCAGCAGCGTCCAATACACGTTTTATTGTGCGTGTTCGTTA 2040

A S M D N N K L N L V L N G A Q I A S L

TGCTAGTATGGACAATAATAAATTGAATCTTGTTTTAAATGGCGCTCAGATAGCATCACT 2100

N V E R T V Q N G G E S L T A L Q C E D

AAATGTGGAACGCACAGTGCAAAATGGCGGCGAATCATTAACAGCTCTTCAATGTGAAGA 2160

F K Y A T F A G D F Q M G S Q S I F G I

TTTTAAATATGCTACATTTGCAGGTGATTTCCAAATGGGTTCTCAATCTATATTCGGTAT 2220

F K D I S N A D F I L D K I E L I P S H

TTTTAAAGACATATCTAATGCAGACTTTATTTTAGATAAAATTGAATTAATCCCATCCCA 2280

F M S S L E Q T Q D D Y S Y N Q N T I Y

TTTCATGTCATCATTAGAGCAAACACAAGATGATTACAGCTATAATCAAAATACTATTTA 2340

T C N Q G Y G T Y D H N S S N M Y D H Q

TACATGTAATCAAGGATATGGTACTTACGACCATAATTCTAGTAATATGTATGATCATCA 2400

N Y K N Y T Q D M D T T Y Q P D Y D N Y

AAACTACAAAAATTATACTCAAGACATGGATACTACATACCAACCAGACTATGATAATTA 2460

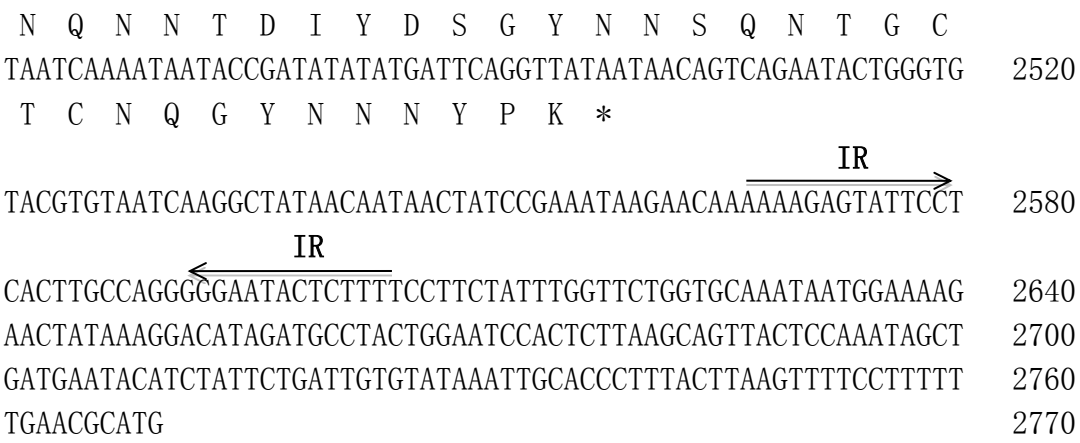


Fig. 28. Nucleotide sequence and deduced amino acid sequence of the *cry20Bb1* gene. The *cry20Bb1* gene is 2,178 bp in length and codes for a polypeptide of 722 amino acids. The potential -35 and -10 boxes and a putative ribosome-binding site (RBS) are marked. The stop codon is marked with asterisks. Five conserved sequence blocks (blocks 1 to 5) are shadowed. Terminal inverted repeats (IR) are indicated below the arrow. The N-termini of 50 kDa (LLVHV) and 30 kDa (NVNIQ) degradation products are double underlined.

Translation map - *cry27Ab1*

```
CCAGCTACAAAAATGATTGTTGGAATAGAGGCCATGCATATAGTCAAAAAAGGTCAACTA     60
AAATTAAGGGTACAACCTGACAAAAATCAGAATAGATGTATTCATTAGTTATTTGGATTA    120
ACTGCTTAAGAGTGGATCCCGTTAGGAATCTATGCCTTTTGTAGCCTTTGTCCATTATTT    180
GCAACAGAACCCCTTTATCCTAACCACCATACTCACATCCAAAAATAATGGGTTTATTTG    240
            -35                 -10
TAGAAAGATTGTTACAGGAATACATAGGTACATTACGAATTTTTAAGAAAGACACCTACT    300
                                RBS       M  N  P  Y  Q  D  K
ATATTTATAGAGGTGGCATAAAGACTAAGGGAGGAACCAAAATGAATCCTTATCAGGATA    360
  N  E  Y  K  I  L  D  A  K  R  N  T  C  H  M  S  N  C  Y  P
AGAATGAATATAAAATCTTAGATGCTAAACGAAATACTTGTCACATGTCTAATTGTTATC    420
  K  Y  P  L  A  N  D  P  Q  M  Y  L  R  N  T  H  Y  K  D  W
CCAAATACCCATTGGCAAATGATCCTCAAATGTATTTGCGCAATACTCATTATAAGGATT    480
  L  T  M  C  N  N  T  K  L  A  G  W  I  P  P  G  S  F  E  F
GGCTAACTATGTGTAATAATACCAAGCTTGCAGGTTGGATACCGCCAGGGAGCTTTGAGT    540
  T  W  L  N  A  T  V  A  A  L  T  I  I  S  V  T  T  A  L  F
TTACCTGGCTAAATGCAACTGTCGCTGCACTTACTATCATTAGTGTAACTACAGCTTTAT    600
  I  A  P  P  L  L  V  G  G  V  I  A  A  G  A  A  I  L  A  G
TTATAGCTCCACCTCTTCTGGTAGGAGGTGTTATAGCTGCAGGAGCTGCTATTTTAGCAG    660
  T  L  P  L  L  W  P  A  D  S  K  P  E  D  N  T  F  N  E  I
GTACATTACCTCTTTTATGGCCTGCGGATTCTAAACCTGAAGATAATACATTTAATGAAA    720
  M  N  A  T  E  V  L  L  N  T  K  I  S  D  F  V  R  Q  T  A
TTATGAATGCAACAGAAGTTCTACTTAATACAAAAATATCTGACTTTGTTAGACAAACAG    780
  D  T  K  I  T  S  L  Q  S  L  M  F  Y  Y  N  N  A  L  D  N
CAGATACCAAAATTACTAGTTTACAAAGTTTAATGTTTTATTATAACAACGCTTTAGATA    840
  W  K  K  N  P  N  D  S  A  A  I  N  T  V  S  T  R  F  Q  I
ATTGGAAGAAAAACCCGAATGATTCAGCCGCGATAAATACGGTAAGTACTAGGTTTCAAA    900
  V  N  A  F  F  V  E  A  M  T  A  L  S  M  P  G  Y  E  L  A
TTGTGAATGCTTTTTTTGTCGAAGCTATGACAGCCCTTTCTATGCCAGGATATGAATTAG    960
  Q  L  G  A  Y  A  Q  A  A  N  L  H  L  L  L  L  R  D  G  I
CACAATTAGGTGCATATGCACAAGCAGCTAATCTGCATTTATTACTTTTACGAGATGGAA   1020
  L  Y  A  D  K  W  N  L  A  K  E  A  T  Y  K  Q  G  D  L  H
TTTTATATGCAGATAAATGGAATTTAGCCAAAGAAGCAACCTATAAACAAGGAGACTTAC   1080
  Y  Q  E  F  L  N  Y  R  N  Q  Y  I  N  H  C  S  T  W  Y  T
ATTATCAAGAATTTCTAAATTATAGAAATCAATATATTAACCATTGTTCAACTTGGTATA   1140
  E  G  Q  I  E  A  N  N  K  G  N  G  L  V  Y  Q  R  T  M  T
CTGAAGGACAAATAGAGGCAAATAATAAAGGTAATGGACTTGTATATCAAAGAACTATGA   1200
  I  L  V  L  D  L  I  A  M  F  S  T  Y  D  P  R  L  Y  T  M
CAATTTTAGTACTAGATTTAATTGCAATGTTTTCAACATACGATCCACGCTTATATACGA   1260
```

P T K T E I L T R T L Y T D G V N R N Q
TGCCAACTAAAACCGAAATTTTAACAAGAACACTTTATACAGATGGAGTAAATAGAAATC 1320
T R S I H N P G L F R R L E Q M E L H T
AAACTAGATCTATACACAATCCAGGTTTATTCCGGCGGTTAGAACAAATGGAATTACACA 1380
Y E Y Q G A Q F L S G H Q N K F R S M N
CTTATGAATATCAGGGTGCACAGTTTTTAAGTGGGCACCAAAATAAATTTAGAAGCATGA 1440
Y N H P L I D G P V Q G Y S S S N I N K
ATTATAATCATCCTCTAATTGACGGTCCCGTACAAGGGTATAGTTCATCGAATATAAATA 1500
I T T I N L G D Y D K I Y S I K T E S R
AAATAACAACTATTAATCTAGGTGATTATGATAAAATTTATAGTATTAAAACAGAAAGCA 1560
D R I V Q G S I T F D K I N F Y G A F N
GAGATCGTATAGTTCAGGGCTCAATTACATTTGATAAAATTAATTTCTATGGGGCATTTA 1620
K S W L F S V Y N Q N G P I I K H S N I
ATAAAAGTTGGCTATTTTCTGTATACAATCAGAACGGTCCAATTATAAAACACAGCAATA 1680
P G V E A P S A T L D Y R N Y T H Y L S
TACCAGGTGTTGAGGCCCCTTCAGCAACGTTAGATTACAGAAACTATACTCATTATTTAT 1740
N C I F Q S N R N G V S E P D Y N T Q S
CAAATTGTATCTTCCAATCAAACCGAAATGGAGTATCTGAACCAGATTACAACACCCAAT 1800
Y I F G W N H N T I D P T G N Y V T D A
CATATATATTTGGCTGGAATCACAATACTATTGATCCCACAGGGAATTATGTAACAGATG 1860
S F V D N G L P E G R Y V P Q I S Q V P
CAAGTTTCGTAGATAACGGCTTACCTGAAGGACGATATGTACCCCAAATTTCACAAGTGC 1920
A V K A S D I Y N P G R V V N A T V E V
CCGCTGTAAAAGCTAGTGATATATACAACCCAGGTCGTGTAGTTAATGCAACAGTTGAAG 1980
G P Y F T G G D V I V S K A Q L D G S G
TTGGACCATATTTTACAGGTGGCGATGTTATTGTATCGAAAGCTCAATTAGATGGATCAG 2040
L A R T V I T F P I I P K R Y Q A S G F
GTCTAGCCAGAACAGTTATAACATTCCCTATTATACCGAAAAGGTATCAAGCAAGCGGAT 2100
R V R M Y Y A A N H I G Q L S Y R A K D
TTCGTGTACGTATGTACTATGCTGCCAATCATATTGGTCAATTGAGTTACCGTGCGAAAG 2160
L N V T G Y A N F T K T F D G W E Y F R
ATCTAAATGTAACCGGTTATGCAAATTTTACAAAAACATTTGATGGCTGGGAATATTTTA 2220
A R H E H F K Y I E F D T T F S L R N S
GAGCGCGACACGAACATTTTAAATATATAGAATTTGATACGACATTTAGCTTACGAAATT 2280
G Q L E E H E L Q I Y Y P N T S R V S G
CAGGTCAATTAGAGGAACATGAATTACAAATTTATTATCCCAATACTTCAAGAGTATCTG 2340
D Q L L I I D K I E F I P V G I R L N Q
GCGATCAATTATTAATTATAGACAAAATTGAATTTATACCGGTGGGAATTCGACTAAATC 2400
P S E G Y N T Y D Q N T N S Y N Q N Y N
AACCATCAGAAGGATATAATACGTACGATCAGAATACTAATAGCTATAATCAAAACTATA 2460

```
  N  Y  N  Q  N  M  D  T  T  Y  Q  P  N  Y  D  N  Y  N  Q  N
ACAATTACAACCAAAACATGGATACTACATATCAACCAAACTATGACAATTATAATCAGA    2520
  S  S  G  M  Y  D  N  P  Y  N  Q  N  P  K  D  S  Y  N  Q  N
ATTCCAGTGGTATGTATGACAATCCCTATAACCAAAATCCCAAGGATAGTTATAATCAAA    2580
  Y  T  D  T  Y  D  S  G  Y  N  N  S  Q  N  V  G  S  N  Y  N
ATTATACCGATACGTATGATTCAGGATATAATAACAGCCAGAACGTTGGCTCTAATTATA    2640
  Q  E  Y  N  T  Y  N  Q  D  T  E  N  M  Y  N  Q  S  Y  N  N
ATCAAGAATACAATACTTATAACCAAGATACGGAAAATATGTATAACCAATCGTATAACA    2700
  Y  N  S  D  N  N  N  Y  N  Q  N  S  D  C  V  C  S  P  G  Y
ATTATAATTCTGACAACAACAATTACAACCAGAATTCTGATTGCGTGTGTAGTCCAGGAT    2760
  N  G  N  Y  E  C  R  C  N  Q  R  A  N  G  N  Y  P  K  *
ATAATGGTAACTATGAATGTAGATGTAACCAAAGGGCTAACGGTAATTATCCTAAATAAG    2820
                                IR                IR
AACAGTAAAAACATACTAAGATTAGTAGTGCGGACCAAGGCATTGGTTCGCACTACTATT    2880
TTTTATATAGTAGGGGTCACCATACATGCCCATCAACTTAAGGATGGATACAAAACAAAA    2940
TTTTGTTCATTTGAACTGAAATCTCTCCAAATGGGGGTAAACCCAGCATCTTCGAATGAA    3000
CCATTAATTTCGTAATTTTCTAATAATTAATAGGTATTTTCATGTTTTTTGGGCTTGAAA    3060
TCCAAGTG                                                        3068
```

Fig. 29. Nucleotide sequence and deduced amino acid sequence of the *cry27Ab1* gene. The *cry27Ab1* gene is 2,478 bp in length and codes for a polypeptide of 825 amino acids. The potential -35 and -10 boxes and a putative ribosome-binding site (RBS) are marked. The stop codon is marked with asterisks. Five conserved sequence blocks (blocks 1 to 5) are shadowed. Terminal inverted repeats (IR) are indicated below the arrow.

Translation map – *cry4Aa*

```
                                                  -35
GAATCGCGCATATTTCAGGTGAATTTTATTTCGAATATGAATAATACATATTTTCCTTAC     60
-10
TTATACTCCTATCCCTACCCCAAAAAAGAATCTATGCTCAGATTCTTTTTTATCTTTTGT     120
CTACAGGAATAGTCAATCTTTTATCCAAATAAATATAAGATTGGAAATTAAATTGAAAGT     180
RBS          M  N  H  Y  N  N  E  N  Y  E  I  I  D  S  N  T
GGAGGAACATATATGAATCACTATAACAATGAAAACTATGAAATTATCGATTCCAATACT     240
 S  P  Y  P  S  N  R  N  N  P  Y  S  R  Y  P  Y  A  N  N  P
TCCCCGTATCCTTCTAACAGAAATAATCCATATTCTAGATATCCTTATGCAAATAATCCT     300
 N  Q  S  L  Q  N  K  N  Y  K  D  W  M  S  I  P  Q  P  Y  A
AATCAATCATTGCAAAATAAAAATTATAAAGATTGGATGAGTATACCTCAACCTTATGCA     360
 D  C  D  N  N  S  F  D  W  L  A  A  V  S  A  G  V  I  V  I
GATTGTGATAACAATTCATTTGATTGGCTTGCCGCTGTTAGTGCAGGTGTCATTGTAATA     420
 G  T  M  L  A  A  F  A  A  P  I  A  A  P  V  L  A  G  S  I
GGTACTATGTTAGCTGCTTTTGCTGCTCCTATAGCAGCGCCAGTTCTAGCTGGATCTATT     480
 I  I  S  I  G  T  L  L  P  I  L  W  P  L  G  Q  S  D  N  N
ATTATATCAATCGGTACATTACTTCCTATTCTTTGGCCACTTGGTCAATCAGATAATAAC     540
 A  V  W  Q  K  F  L  D  Q  G  N  N  L  T  C  Q  Q  L  T  P
GCAGTATGGCAAAAATTCCTTGATCAAGGAAATAACCTTACATGTCAACAACTAACTCCA     600
 G  I  K  V  A  V  D  A  A  L  N  N  L  R  V  Q  A  H  Y  F
GGAATTAAAGTAGCAGTAGATGCAGCTTTAAATAATTTAAGAGTTCAAGCCCATTATTTC     660
 N  D  A  V  T  Y  W  E  K  S  I  G  T  S  N  E  I  D  A  R
AACGATGCCGTTACCTATTGGGAAAAAAGTATAGGCACTTCAAATGAAATAGATGCTAGA     720
 N  N  A  R  D  I  Y  I  N  A  V  Q  I  I  E  G  L  M  P  I
AATAATGCAAGAGATATTTATATAAATGCTGTACAAATAATTGAAGGACTTATGCCTATA     780
 F  K  S  S  G  Y  E  V  L  L  L  S  T  Y  A  Q  A  A  L  L
TTTAAATCATCAGGTTATGAAGTATTATTATTATCTACTTATGCACAAGCCGCATTATTA     840
Q  V  T  L  L  H  Q  G  I  Q  Y  A  S  K  W  N  L  A  R  D
CAAGTTACTTTACTACATCAAGGCATTCAATATGCTTCCAAATGGAATTTAGCTCGAGAT     900
T  G  D  F  Y  R  Q  R  L  Y  E  A  I  D  R  H  I  D  Y  C
ACTGGAGATTTTTATCGTCAAAGACTTTACGAAGCAATAGATAGACACATTGATTATTGT     960
E  T  W  Y  Q  T  G  L  D  E  L  K  K  N  E  N  L  T  F  A
GAAACATGGTATCAAACAGGTCTAGACGAACTCAAGAAAAACGAAAATTTAACATTTGCT    1020
A  Y  I  N  Y  R  R  E  Y  T  I  N  V  L  D  V  I  S  L  I
GCCTATATAAATTATCGTAGAGAATATACTATCAATGTATTAGATGTTATTTCCCTAATT    1080
P  A  L  D  L  R  I  Y  P  D  T  K  P  I  N  I  E  F  T  R
CCAGCATTAGATTTACGTATTTATCCAGACACTAAACCAATTAACATAGAATTCACGCGA    1140
N  I  F  T  A  I  P  T  S  N  Q  S  R  I  S  A  F  I  G  R
```

AATATATTTACAGCTATACCAACTTCAAACCAATCAAGAATAAGTGCATTTATAGGACGT 1200
E N I E K L E K E L W P S T E L F T Q L
GAAAACATTGAAAAATTAGAAAAAGAACTATGGCCTTCTACAGAATTATTTACACAGTTA 1260
R Q I A F Y Q D Y N Y I E S G N Y L A Q
AGACAAATAGCTTTTTATCAAGATTATAATTATATCGAAAGCGGAAACTACTTAGCTCAG 1320
I T N L I V H A N D S N L I T K T Y G N
ATAACTAATCTAATCGTTCACGCTAATGATTCTAACCTCATTACAAAAACATATGGAAAC 1380
A P S Q T A S P T I V L S P Q Q S I Y N
GCACCATCACAAACAGCATCTCCAACTATAGTTCTCTCCCCACAACAAAGCATTTATAAT 1440
C T I D N Y S I G D T P R T G I K Y M E
TGTACAATAGATAATTATTCAATTGGAGATACACCAAGAACAGGAATAAAATATATGGAG 1500
L K V A T Q R I T S N S V K F G S E T S
CTAAAAGTAGCTACTCAAAGAATTACCAGTAATTCTGTAAAATTCGGTTCGGAAACTAGT 1560
G S Y Q R N I L P F P T D I Q V T S L Q
GGGTCTTACCAAAGGAATATATTACCGTTTCCAACAGATATACAAGTAACAAGTTTACAA 1620
N Y Q Y K L S R I T M S Q N K Y P M T S
AACTATCAATATAAATTATCTCGTATCACAATGTCTCAAAATAAATATCCCATGACTTCG 1680
V G E T T T Y L Y G F I W T H A Q S N P
GTCGGCGAGACTACTACTTATCTATATGGATTTATTTGGACACACGCTCAATCTAATCCT 1740
T N T I T S K N K N N Q K T I T Q I S A
ACAAATACCATTACTTCTAAAAATAAAAATAATCAAAAAACAATTACACAAATCTCTGCA 1800
V K A Y E L S N P N S H I F P N T I T V
GTTAAAGCATATGAACTTTCCAATCCTAATTCTCATATATTTCCCAATACAATTACAGTT 1860
I E G P G H T G G K L V K S T Y I L D Q
ATAGAAGGACCTGGTCATACAGGTGGAAAACTAGTAAAATCCACCTATATTCTCGACCAA 1920
L V I K C T F T D S S Q Y R L R I R Y A
TTAGTAATTAAGTGTACATTTACTGATAGTAGCCAATATCGATTGCGTATTAGATACGCT 1980
T D I I N N G I L K V T I Q S S N N L N
ACAGATATAATAAATAATGGAATCTTAAAAGTAACTATACAATCTTCAAATAACTTAAAT 2040
I I K R Y E F S L R K G N L N T S T N I
ATTATTAAAAGATATGAATTTAGTCTAAGAAAAGGAAACTTAAATACATCTACAAATATC 2100
P L Y K D F L T T E A L N P F S V T A N
CCCTTATATAAAGATTTCTTAACAACGGAAGCTCTCAATCCATTTTCAGTTACAGCAAAT 2160
E K V N I I I E N G S T N S G T I L I D
GAAAAGGTAAATATAATAATAGAAAATGGATCAACAAATTCAGGAACTATTCTTATTGAT 2220
K L E F V P Q * IR IR
AAACTTGAATTTGTCCCACAATAAACAACAAAAATGAAAGTTCTTTTTCTTTTTTCGGTT 2280
TACAAAATATAGATATTACTTTTAATATCTATATTTATTTGAATTTACATTTATAGAGGG 2340
AAATGTTTTCGTCATACCTTGATCTACAAACACATTCCATGTTGCCCCAATAACATTTGT 2400
ATCATTTAAACGTTTTAATTTAAATTCATATTTCATTAAATATTTTACATATCTAGTTGG 2460

CTCCGCACC 2469

Fig. 30. Nucleotide sequence and deduced amino acid sequence of the *cry4Aa* gene. The *cry4Aa* gene is 2,052 bp in length and codes for a polypeptide of 684 amino acids. The potential -35 and -10 boxes and a putative ribosome-binding site (RBS) are overlined. The stop codon is marked with asterisks. Five conserved sequence blocks (blocks 1 to 5) are shadowed. Terminal inverted repeats (IR) are indicated below the arrow.

Translation map – *cry56Ba1+cry39orf2*

```
CCATTGGTTATTGAACATAGAATGATGAACAGAGTTTTATTACATATAAGACAATAAAAA    60
ACGCTTCTTTTAAATCAAAAAGAAGCGTGTTCAGAAAAGGGAAATTTTCTGTACACAGGT   120
                                                      -35
AAAATTCCCATAATTCGGTTTTCCCTAGAAAAAATAAAATGAATCCAATACATGTTCGGA   180
                -10
TTTATTTTCTACAATTTCTTTCAATGACGAATAATTAAATATCTACAACAATCTATTTAG   240
TATCCCTCTCTTTTTTCGAGGGATAGGAAGAAACATCCAAGTATGAATTTTGTTCATATA   300
AAAGTGAATAACTCTTTCCGCACCCTTAAAAACAACAAAGAAAAAATCGTTCTACAGAAA   360
TCTGAAGCTTTTAAAAAAATACATGCAATACATAAAGAGAAGATTTGAAAATTAAATACC   420
TGACCAAATATAATGGGTTTATTTGTAGAAACATCGTTACAGGAATACATTGGGGTACTT   480
CGAATATATAGAAAGACACCTAGCATATATTTATTAGGTGTTTTAAAAATAAGGACTACA   540
   RBS          M  N  S  Y  Q  N  K  N  E  Y  E  I  L  D  T
TAAGGAGTGAAAAATATGAATTCATATCAAAATAAAAATGAATATGAAATATTGGATACT   600
S  P  N  S  S  T  M  S  T  R  Y  P  R  Y  P  L  A  K  N  P
TCACCAAACAGCTCTACTATGTCTACTCGTTATCCTAGATACCCACTAGCAAAAAATCCA   660
Q  I  S  M  Q  N  T  N  Y  K  D  W  I  N  M  C  T  N  N  T
CAAATATCCATGCAAAATACGAATTATAAAGACTGGATAAATATGTGTACAAATAATACC   720
L  I  P  I  E  P  L  D  L  T  W  Q  N  A  L  V  S  V  F  G
CTTATTCCTATAGAACCTTTAGACCTTACCTGGCAAAATGCTCTTGTTTCAGTCTTCGGT   780
I  A  S  A  V  A  A  L  L  A  A  P  I  T  G  G  T  S  I  A
ATCGCTTCAGCTGTTGCAGCATTGTTAGCAGCTCCAATTACTGGCGGAACATCTATAGCA   840
A  G  A  A  I  I  A  N  I  L  P  L  T  F  P  A  N  A  E  S
GCTGGAGCGGCTATAATAGCTAATATATTACCATTAACCTTTCCCGCTAATGCTGAGAGT   900
V  P  N  K  L  M  D  A  T  Q  E  L  L  G  P  L  E  E  Y  T
GTTCCGAATAAGCTTATGGATGCCACACAAGAATTACTTGGCCCTCTAGAAGAATACACT   960
R  N  R  A  N  S  E  L  L  S  L  S  S  Q  L  E  A  F  K  G
AGAAATAGAGCAAATTCGGAGCTACTCAGTTTGAGTTCACAGTTAGAAGCTTTTAAAGGT  1020
L  F  D  Y  W  L  A  D  R  Q  N  P  N  A  T  N  S  V  S  A
CTATTTGATTATTGGCTCGCTGACCGCCAAAATCCAAATGCAACTAATTCAGTTAGTGCT  1080
R  F  T  A  I  H  N  N  F  I  G  A  M  A  L  F  K  I  P  G
CGTTTTACTGCAATTCATAATAATTTTATAGGGGCAATGGCTCTTTTTAAAATACCGGGT  1140
Y  E  A  L  L  L  P  V  Y  A  Q  A  A  R  L  H  L  L  H  L
TATGAAGCCTTACTGTTACCGGTATATGCTCAGGCTGCACGTTTACATTTGCTTCATTTA  1200
R  D  G  I  T  Y  A  D  Q  W  Q  L  A  D  P  T  N  A  A  Y
AGAGACGGTATCACGTACGCTGATCAATGGCAGTTAGCTGATCCAACTAATGCAGCTTAT  1260
A  G  D  Y  H  Y  S  E  F  K  K  Y  S  A  Q  Y  A  D  H  C
GCGGGAGATTACCACTATAGTGAATTTAAGAAATATTCTGCACAATATGCAGATCATTGT  1320
E  L  V  V  N  N  Q  L  N  K  I  K  N  T  N  G  K  T  W  K
```

GAATTAGTAGTTAATAATCAACTAAATAAGATAAAAAATACAAACGGTAAAACATGGAAA 1380
D Y N E Y R R K M I L S V F D I V A E F
GACTACAACGAATATCGTCGAAAGATGATATTATCTGTTTTCGATATTGTTGCTGAATTT 1440
S T F D P I L Y K G A I N R E I L T R K
TCAACCTTTGATCCAATTTTATATAAAGGAGCGATAAATAGAGAGATTTTAACACGTAAA 1500
I Y T D P V N F T P G F S I A D D E N R
ATATATACAGACCCAGTTAATTTCACACCTGGTTTTTCAATTGCTGATGATGAAAATAGA 1560
Y T V R P S N V K Q L V A S T L F T N V
TATACAGTTAGACCGTCAAATGTTAAACAATTAGTCGCCTCCACACTATTTACTAACGTG 1620
A S A Q Y A G F I G N R N R Y L S L L G
GCATCTGCTCAGTATGCTGGATTTATTGGAAATAGAAATCGTTATTTAAGTTTATTAGGT 1680
G E P L E G P V I G K S V S E N V V A G
GGAGAGCCACTTGAAGGACCTGTAATCGGAAAATCAGTATCCGAAAATGTTGTAGCAGGT 1740
V P T N E S I Y E V G V N G Y P N D Y P
GTACCAACAAATGAATCGATTTATGAAGTTGGTGTAAATGGTTACCCGAATGATTATCCA 1800
R N I G L R W G S L T R F Q N Y Y A G S
CGTAATATAGGTTTGAGATGGGGTTCATTAACTAGATTTCAAAATTATTATGCTGGAAGC 1860
Q Y N L G G L T T V S V P P K N N A P I
CAGTATAATTTAGGGGGGTTAACTACGGTCTCTGTGCCACCTAAAAATAATGCCCCAATA 1920
N N T N F T H R L S D I I L P G N S G S
AATAATACTAATTTTACTCATCGATTATCAGATATAATTCTTCCTGGAAATAGTGGCTCA 1980
S F A W T H V E V N P T E N Y L S T D Q
TCTTTTGCATGGACTCATGTTGAGGTCAATCCTACAGAAAACTATTTATCAACAGATCAA 2040
I N L I S A T K T S T Y N S M W K G P G
ATTAATTTAATATCTGCTACAAAAACTTCAACATATAACAGTATGTGGAAGGGACCTGGA 2100
F I G G D L T S S D I A F G E Y L F Y N
TTTATAGGAGGAGATTTAACAAGCAGCGACATAGCATTTGGAGAATACTTATTTTATAAT 2160
F K Y K S P G S S A R F K I R L R Y G S
TTTAAGTACAAATCCCCTGGTAGCTCAGCTAGGTTTAAAATTCGTTTACGTTATGGATCT 2220
W G S Y G S V Y Y I L G N T T S P K T L
TGGGGTAGTTATGGGTCGGTATACTATATATTAGGGAATACCACTTCACCAAAGACTCTT 2280
F E N T R L D L N N Y K Y D Q F K V V E
TTTGAAAATACTAGATTAGATCTTAATAATTATAAGTATGATCAATTTAAAGTAGTAGAG 2340
L W G T A E N I T D N N L I I K V A F A
CTTTGGGGAACTGCAGAAAATATTACAGACAACAATTTAATTATTAAAGTAGCCTTTGCT 2400
N T G G S T G F Y L D R L E L I P M T G
AACACAGGAGGTAGTACTGGGTTTTATCTAGATAGATTGGAATTAATCCCTATGACAGGG 2460
M P T E Y T E P Q K L E T A Q K A V N D
ATGCCAACAGAATACACTGAACCGCAAAAATTGGAAACAGCACAGAAAGCAGTAAACGAT 2520
L F T N *

TTATTTACCAATTAAAAAAAAGTATGTAATGAAGTAGGTAGTAATCCTGTTCAAAAAATA 2580

RBS M Y T N A M K N T L K I E

CGCAGAAAAGGTAGTGAATCCTATGTATACCAATGCTATGAAAAATACATTAAAAATAGA 2640

T T D Y E I D Q A A I S I E C M S H E K

AACGACGGATTATGAAATAGATCAAGCGGCCATTTCTATAGAATGTATGTCACATGAAAA 2700

Y P Q E K M I L W D E V K Q A K Q L S Q

ATATCCGCAAGAAAAAATGATATTATGGGATGAAGTAAAACAAGCAAAACAACTCAGTCA 2760

S R N L L Y N G D F E D A S N G W K T S

ATCTCGTAATTTACTCTACAATGGGGATTTTGAAGATGCATCAAACGGATGGAAAACAAG 2820

Y T I E I R K N S P I F K G Q Y L H M F

TTATACGATTGAAATTCGAAAGAATAGTCCCATTTTTAAAGGGCAGTACCTTCATATGTT 2880

G A R D V L G E V F P T Y V Y Q K I D E

TGGTGCAAGAGATGTTTTAGGTGAAGTGTTTCCAACATATGTGTATCAAAAAATTGATGA 2940

S K L K P Y T R Y R V R G F V G S S K D

GTCTAAATTAAAACCATATACACGTTATCGAGTAAGAGGATTTGTGGGAAGTAGTAAAGA 3000

L K L A V T R Y G K E I D A I M D V P N

TCTAAAACTAGCGGTAACACGTTACGGGAAAGAAATTGATGCCATTATGGATGTTCCAAA 3060

D L A Y M Q P N P S C G D Y R C D S P S

TGATTTGGCCTATATGCAGCCTAACCCTTCATGTGGAGATTATCGCTGTGACTCACCATC 3120

Q S M M S H G Y P T P V T D G S A S N M

CCAGTCTATGATGAGTCACGGATATCCTACACCAGTAACAGATGGATCTGCTTCTAATAT 3180

Y A C P S D R V K K H V K C H D R H L F

GTATGCATGCCCGTCAGACCGAGTTAAAAAACATGTGAAGTGTCACGATCGCCATCTATT 3240

D F H I D T G E L D T N T N L G I L V L

TGATTTTCATATTGACACAGGAGAGTTAGATACAAATACAAACTTAGGTATCTTGGTCTT 3300

F K I S H P N G Y A T L G N L E V I E E

ATTTAAGATTTCCCATCCAAATGGATACGCTACATTAGGGAATCTAGAAGTGATTGAAGA 3360

G P L T D E A L E H V R Q K E K K W N R

AGGGCCACTAACAGACGAAGCATTGGAACATGTGAGACAGAAAGAAAAGAAATGGAATCG 3420

H I E K A R M E T Q Q A Y D P A K Q A V

ACACATAGAGAAAGCGCGAATGGAAACACAACAAGCTTATGATCCAGCAAAACAGGCAGT 3480

D A L F T S A Q E L H Y H T T L N H I K

AGATGCATTATTTACAAGTGCACAAGAGTTACACTATCATACTACTTTAAATCATATTAA 3540

N A D Q L V Q S I P Y V N H A G L P D A

GAATGCCGATCAGTTGGTACAGTCGATTCCCTATGTAAACCATGCTGGGTTACCGGATGC 3600

P G M N Y D L Y Q G L N A R I M Q A Y N

TCCAGGTATGAACTATGATTTATATCAAGGGTTAAACGCGCGTATCATGCAGGCGTACAA 3660

L Y D A R N V I T N G D F T Q G L Q G W

TTTATATGATGCACGAAATGTCATCACAAATGGTGACTTTACACAAGGATTACAGGGATG 3720

H A T G N A A V Q Q M D G A S V L V L S

GCACGCAACAGGAAATGCCGCGGTACAACAAATGGATGGCGCTTCTGTATTAGTTCTATC 3780
N W S A G V S Q N L H A Q D H H G Y V L
AAACTGGAGTGCCGGGGTATCTCAAAATCTGCATGCCCAAGATCATCATGGATATGTGTT 3840
R V I A K K E G T G K G Y V T M M D C N
ACGTGTGATTGCCAAAAAAGAAGGGACCGGAAAAGGGTATGTAACGATGATGGATTGTAA 3900
G K Q E T L T F T S C E E G Y M T K T V
TGGAAAGCAGGAAACACTTACGTTCACTTCTTGTGAAGAAGGATATATGACAAAAACAGT 3960
E V F P E S D R V R I E I G E T E G T F
AGAGGTATTCCCAGAAAGTGATCGTGTACGGATTGAAATAGGAGAAACCGAAGGTACATT 4020
Y I D S I E L L C M K G Y P S N Y N Q N
TTATATAGATAGCATAGAGTTACTTTGTATGAAAGGGTATCCTAGCAATTACAACCAAAA 4080
T D N M Y E Q S Y N G N Y N Q N T S D V
TACAGATAATATGTATGAGCAAAGTTATAATGGAAATTATAATCAGAATACTAGCGATGT 4140
Y H Q G Y T N N Y N K D S S S M Y N Q N
GTATCACCAAGGATATACAAACAACTATAACAAAGACTCTAGTAGTATGTATAATCAAAA 4200
Y T N N D D Q H S G C T C N Q G H N P G
TTATACTAACAATGATGACCAGCATTCCGGCTGCACATGTAACCAAGGGCATAACCCTGG 4260
C T C N Q R Y N R * IR IR
CTGTACATGTAATCAAAGATATAACCGTTAACGATTCTAAATAAGAATTAAAATCATTGC 4320
GAAAAATAAAAACCAACTCACAAAATCTATTGCCTATCATAACATAAGCTTTACAAATAA 4380
CTGACATATTCTAGAAGCGGTCTCCTTAATTCTAAAATAAGGAGATCCTTTTCGTTTCCA 4440
CAATATCGATTAATGAAAATACTCCTTTATAGAACGATTTAGGCTGATTGGATTTGAATG 4500
TTGTTGAATCGATGATGAATCTTGAATAGAAATTTGGTATTCAAACTTCTGACGAGGTAT 4560
CTG 4563

Fig. 31. Nucleotide sequence and deduced amino acid sequence of the *cry56Ba1* and *cry39orf2* gene. The *cry56Ba1* gene is 1,980 bp in length and codes for a polypeptide of 659 amino acids. The *cry39orf2* gene is 1,689 bp in length and codes for a polypeptide of 562 amino acids. The potential -35 and -10 boxes and a putative ribosome-binding site (RBS) are marked. The stop codon is marked with asterisks. Five conserved sequence blocks (blocks 1 to 5) are shadowed. Terminal inverted repeats (IR) are indicated below the arrow.

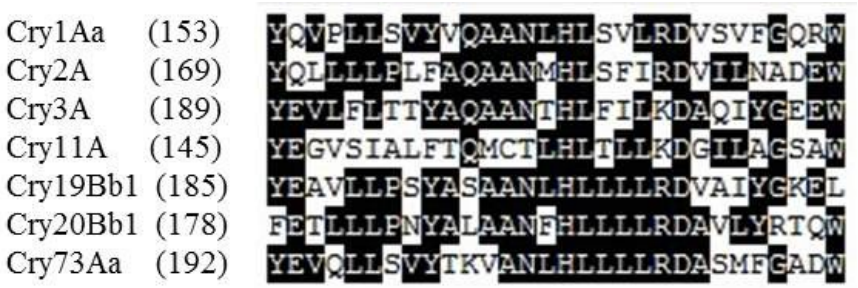
3.2 Analysis of the deduced amino acid sequence of the novel *cry* genes

A Clustal W comparison of the novel proteins from *mogi* with other known Cry protein sequences (http://www.biols.susx.ac.uk/Home/Neil-Crickmore/Bt/) helped to identify the characteristic Cry conserved blocks predicted by Schnepf *et al*. (1998). All of these *cry* genes (*cry19Bb1, cry73Aa, cry20Bb1, cry27Ab1, cry4Aa* and *cry56Ba1*) contained the 5 conserved amino acid residue blocks (Fig. 32 and Fig. 33, block 1 to block 5) that are present in almost all Cry proteins (Schnepf *et al.,* 1998), but does not contain the carboxyl-terminal half of the typical 130 kDa-type crystal proteins (Boonserm *et al.,* 2005; Galitsky *et al.,* 2001; Grochulski *et al*., 1995).

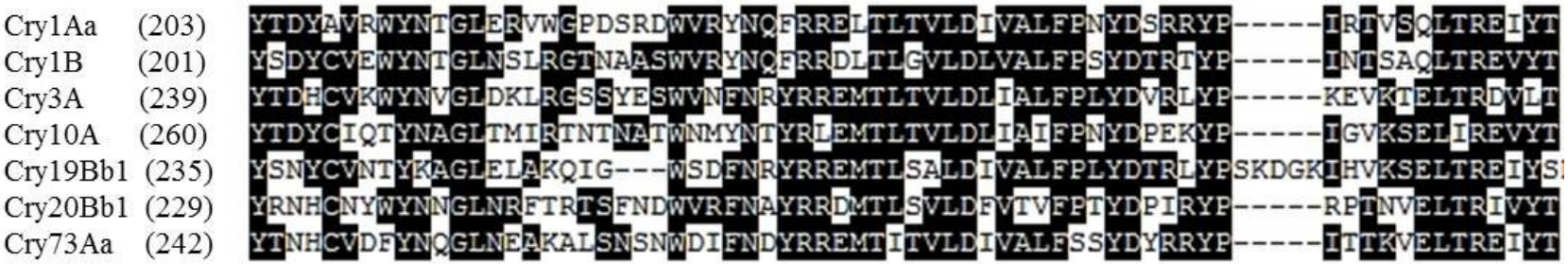
3.3 Transcription level analysis of *cry* genes in *B. thuringiensis* subsp. *mogi*

To confirm the transcription profiles of the selected *cry* genes in *B. thuringiensis* subsp. *mogi* strain, RT-PCR and qPCR were carried out as previous described. The transcription levels of *cry* genes were compared to those of the 16S rRNA gene at each time point. The results suggested that all of these *cry* genes were successfully transcribed in *B. thuringiensis mogi* strain (Fig. 34) in different expression time with different maximum levels (Fig. 35). *B. thuringiensis* subsp. m*ogi* contained *cry27Ab1, cry19Bb1*, *cry20Bb1*, *cry56Ba1*, *cry4Aa* and *cry73Aa* genes, with *cry27Ab1* being transcribed at much higher level and *cry4Aa* being expressed at a relative lower level.

Block 1

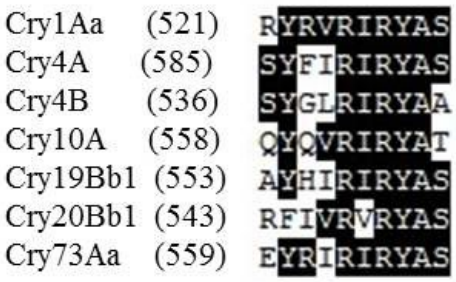


Block 2



Block3

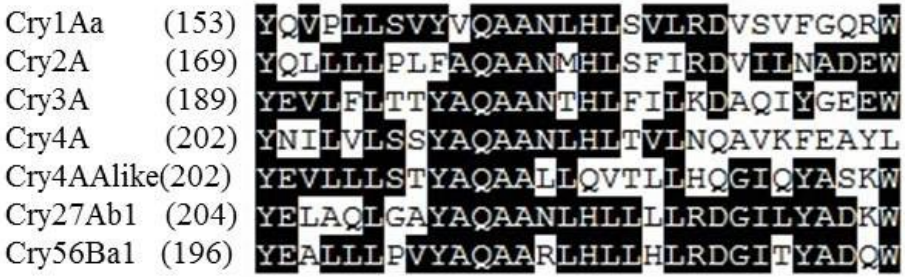
Block 4



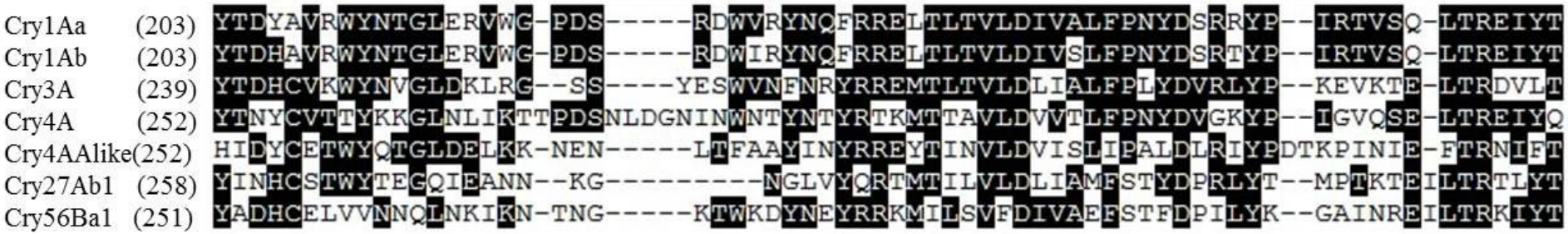
Block 5

Fig. 32. Comparison of the deduced amino acid sequence of Cry19Bb1, Cry20Bb1, Cry73Aa with other Cry proteins in five conserved blocks.

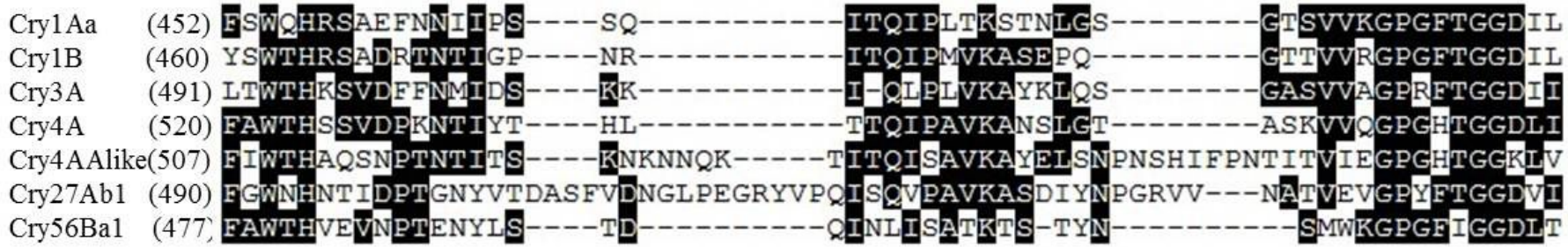
Block 1



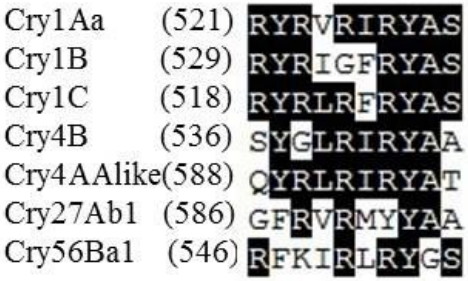
Block 2



Block3



Block 4



Block 5

Fig. 33. Comparison of the deduced amino acid sequence of Cry4Aa, Cry27Ab1, Cry56Ba1 with other Cry proteins in five conserved blocks.

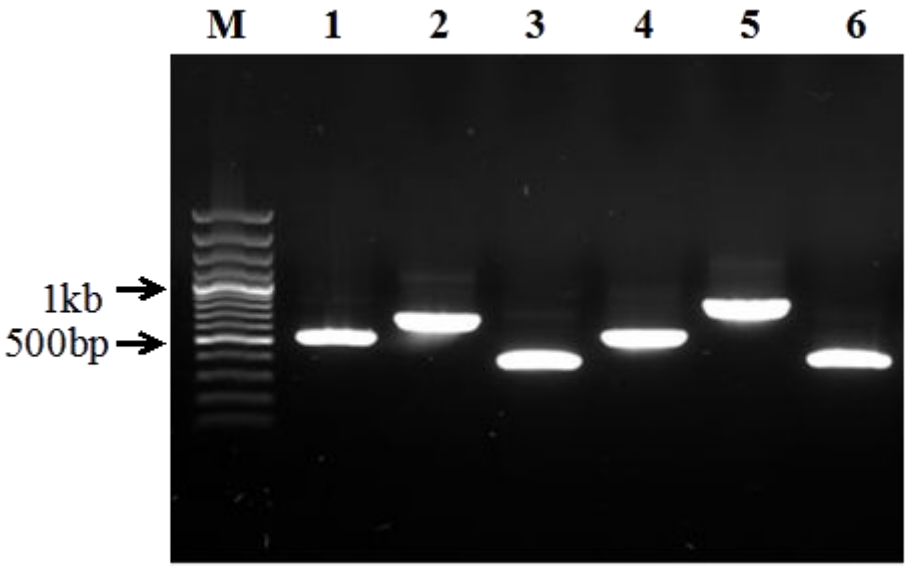


Fig. 34. Agarose gel electrophoresis analysis of *cry* genes mRNA obtained by RT-PCR from *B. thuringiensis* subsp. *mogi* in the sporulation stage. Lane M, Gene Ruler^{TM} 100 bp DNA ladder; lanes 1, *cry27Ab1*; 2, *cry19Bb1*; 3, *cry20Bb1*; 4, *cry56Ba1*; 5, *cry4Aa* and 6, *cry73Aa.*

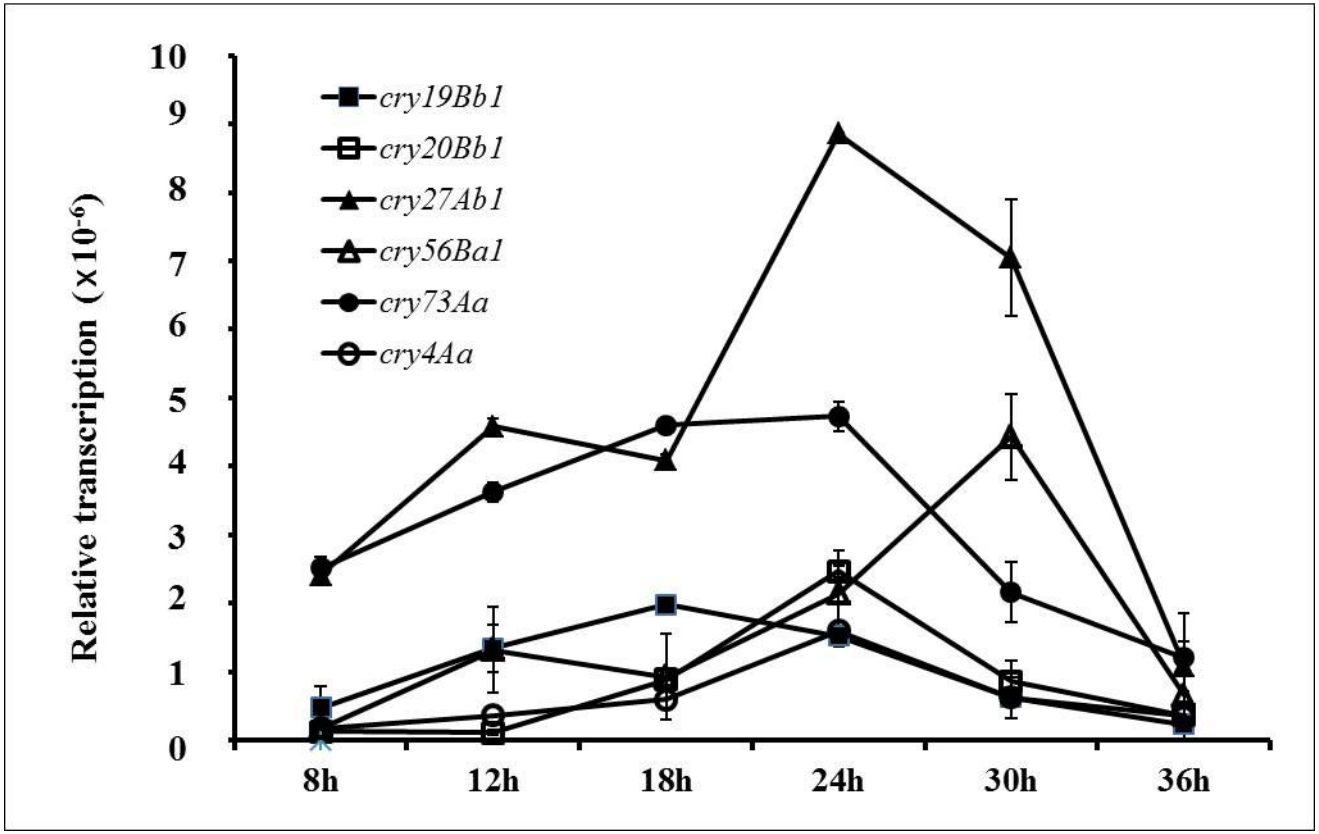


Fig. 35. Transcription-level analysis of the 6 *cry* genes from *B. thuringiensis* subsp. *mogi* were analyzed. The error bars indicate standard deviations.

3.4 Cloning and expression analysis of mosquitocidal *cry* genes in Cry$^-$B

The PCR products (Fig. 36) were cloned into pHT1K (Fig. 22B) and their DNA sequences were confirmed. To investigate the interaction of insecticidal *cry* genes in acrystalliferous *B. thuringiensis*, the recombinant plasmids containing different *cry* genes were introduced into Cry$^-$B by electroporation separately. The expression of these genes were analyzed by SDS-PAGE (Fig. 37).

Cry19Bb1, Cry73Aa with Cry40ORF2, Cry27Ab1 and Cry4Aa, which expressed under their original promoter, were failed to detect on the SDS-PAGE (Fig. 37, lane 2, 3, 5, 6). Meanwhile, there were no crystals visualized in the phase contrast micrographs (Fig. 38, panel 2, 3, 5, 6) either.

The production of Cry20Bb1 by transformant harboring recombinant pHT1K -20Bb1 in Cry$^-$B was shown as the 50 kDa and 30 kDa bands in lane 4 (Fig. 37). The 50- and 30-kDa proteins were confirmed to be the degraded products of the intact 82-kDa Cry20Bb1 protein by N-terminal sequencing. The N-terminal sequences of the 50- and 30-kDa proteins were LLVHV and NVNLQ (Fig. 28), respectively, showing that cleavage occurs between Asp-98 and Leu-99 and between Thr-215 and Asn-216 of the intact 82-kDa Cry20Bb1 protein. This cleavage pattern indicates that both the 50- and 30-kDa proteins have truncated domain I and domain III that are required for insecticidal activity (Chen *et al*., 1993; 1995). The production of Cry56Ba1 with Cry39ORF2 was also detected as the 70 kDa band in lane 7 (Fig. 37).

Transformations crystals of Cry20Bb1 and Cry56Ba1 were visible when phase-contrast microscopy was used, and shown an much smaller inclusion compare with wild type strain (Fig. 38, panel 4 and 7). TEM also demonstrated this results (Fig. 39).

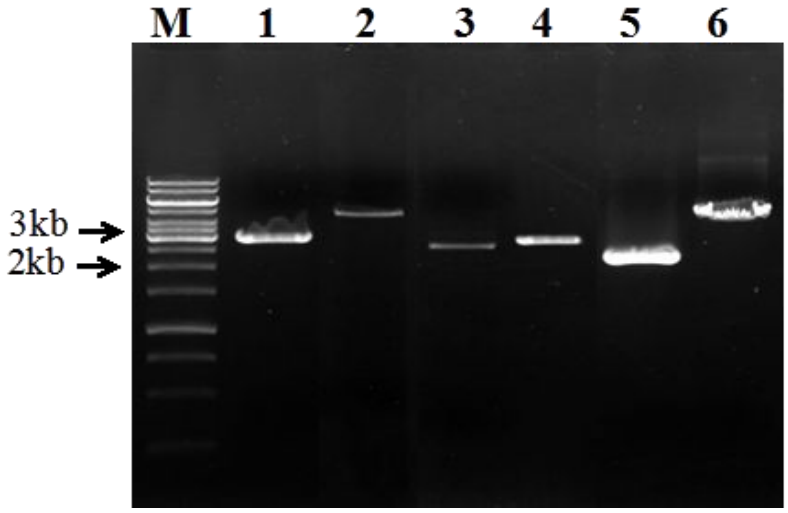


Fig. 36. Agarose gel electrophoresis analysis of PCR products. Lane M, Gene RulerTM 1 kb DNA ladder; lanes 1, *cry19Bb1*; 2, *cry73Aa+cry40orf2*; 3, *cry20Bb1*; 4, *cry27Ab1*; 5, *cry4Aa* and 6, *cry56Ba1+cry39orf2*.

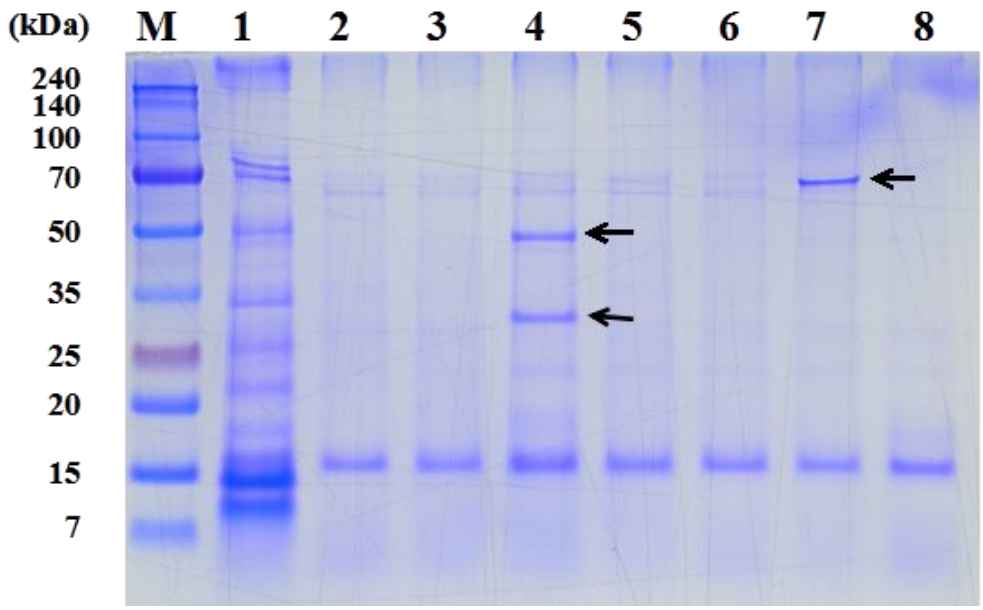


Fig. 37. SDS-PAGE analysis of the recombinant *B. thuringiensis cry⁻B* strains containing different *cry* genes. M, molecular mass marker; lane 1, wild type *mogi* strain; 2, CB/pHT1K-19Bb1; 3, CB/pHT1K-73Aa+40orf2; 4, CB/pHT1K-20Bb1; 5, CB/pHT1K-27Ab1; 6, CB/pHT1K-4Aa; 7, CB/pHT1K-56Ba1+39orf2; 8, wild type Cry⁻B strain.

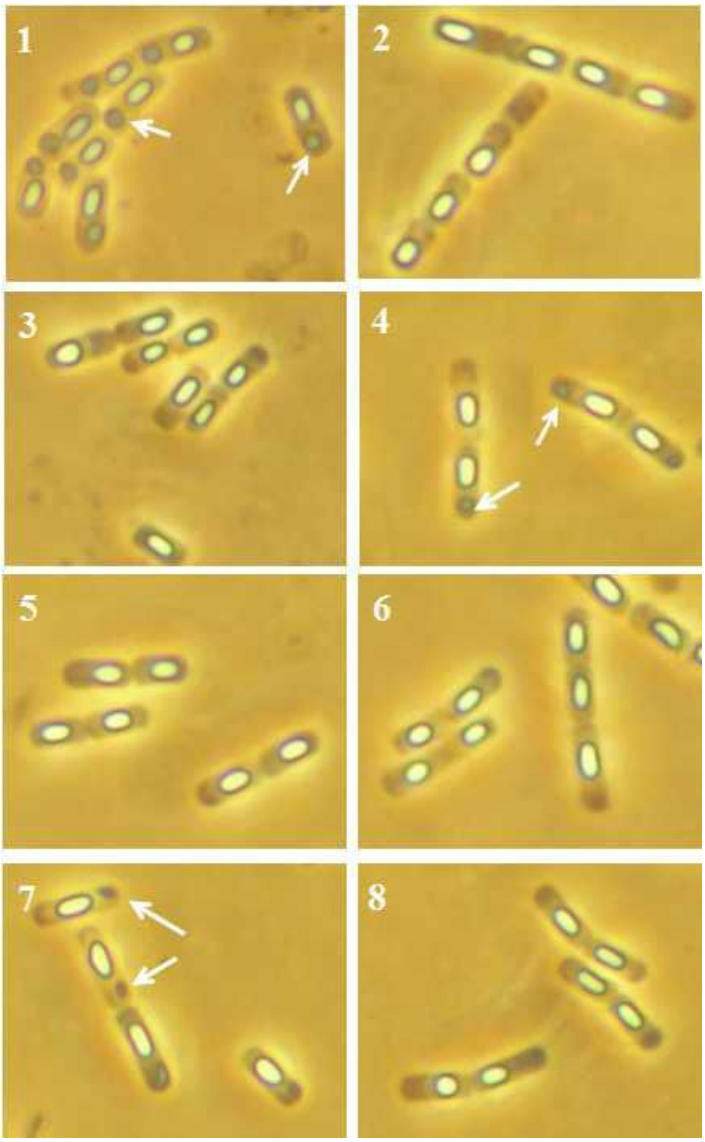


Fig. 38. Phase contrast microscopies of the recombinant *B. thuringiensis* Cry-B strains containing different *cry* genes. 1, wild type *mogi* strain; 2, CB/pHT1K-19Bb1; 3, CB/pHT1K-73Aa+40orf2; 4, CB/pHT1K-20Bb1; 5, CB/pHT1K-27Ab1; 6, CB/pHT1K-4Aa; 7, CB/pHT1K-56Ba1+39orf2; 8, wild type Cry-B strain. Arrows indicate crystal.

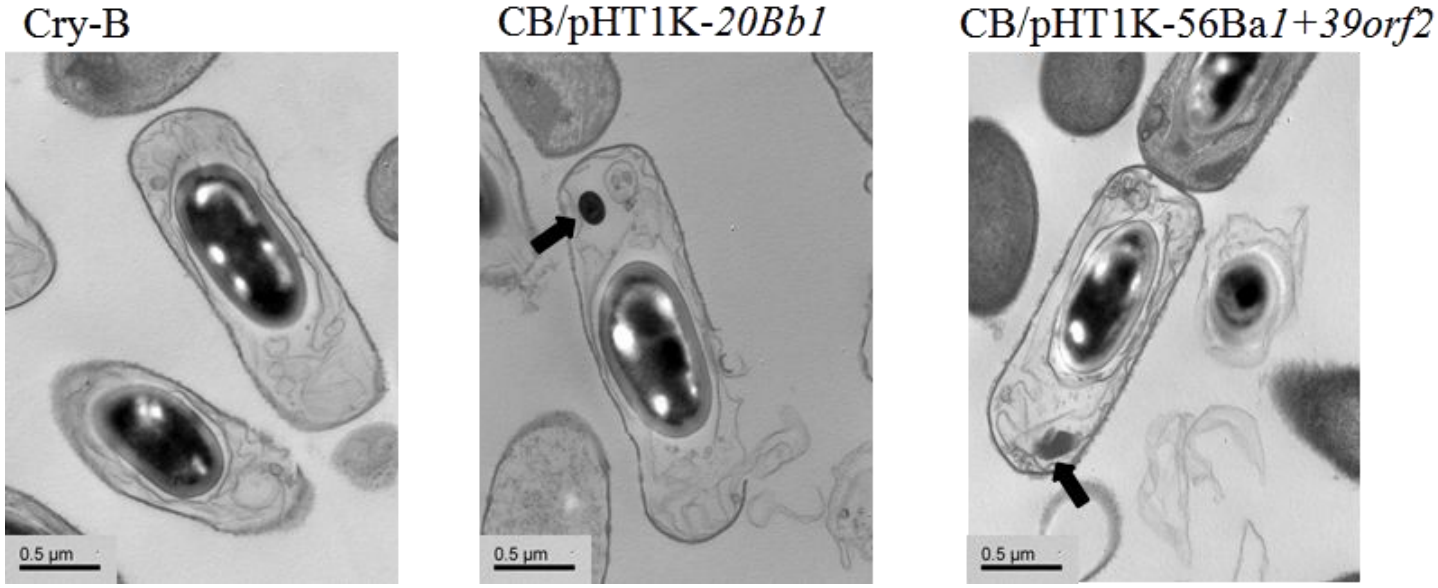


Fig. 39. Transmission electron microscopy of the acrystalliferous *B. thuringiensis* Cry-B strain and the recombinant strains. Magnification is 40,000 ×. Arrows indicate the inclusion.

Table 15. Toxicity of recombinant strains of *B. thuringiensis* Cry-B against *Culex pipiens molestus* and *Cluex pipiens pallens* 4th instar larvae.

Strain	Toxin combination	*Culex pipiens molestus*	*Cluex pipiens pallens*
		LC_{50}[a] (mg/ml)	LC_{50}[a] (mg/ml)
CB/pHT1K-19Bb1	Cry19Bb1	Nt[b]	Nt[b]
CB/pHT1K-73Aa+40orf2	Cry73Aa + Cry40orf2	Nt[b]	Nt[b]
CB/pHT1K-20Bb1	Cry20Bb1	1.17 (0.89-2.06)	1.05 (0.95-1.67)
CB/pHT1K-27Ab1	Cry27Ab1	Nt[b]	Nt[b]
CB/pHT1K-4Aa	Cry4Aa	Nt[b]	Nt[b]
CB/pHT1K-56Ba+39orf2	Cry56Ba1 + Cry39orf2	1.02 (0.91 - 1.54)	1.26 (1.10 -1.42)
Cry-B	no	Nt[b]	Nt[b]

[a]LC_{50}: 50% lethal concentration (in μg) of freeze-dried spore–crystal complex per milliliter after 48 hours. The data are the total of three assays as determined by Probit analysis. [b]Nt, not toxic at 10 mg/ml. [c]FL_{95}: fiducial limits at P=0.95.

3.4 Toxicity of transfromant

The toxicity of wild-type and recombinant strains of *B. thuringiensi* was evaluated against 4th instars of *C. pipiens molestus* and *C. pipiens pallens* (Table 15). Strains CB/pHT1K-19Bb1, CB/pHT1K-73Aa+40orf2, CB/pHT1K-27Ab1 and CB/pHT1K-4Aa, which did not produce visualized inclusions, were not toxic even at 10 mg/ml, while strains CB/pHT1K-20Bb1 and CB/pHT1K-56Ba+39orf2, which produced small but apparently inclusions, showed moderate toxicity, with no significant differences in their median lethal concentrations (LC_{50}).

3.5 Over expression of *cry56Ba1* operon and functional analysis of Cry39ORF2

(i) Cry39ORF2 is required for crystallization of Cry56Ba1

When sporulated cells were examined by phase-contrast microscopy, crystalline inclusions were produced only when both *cry56Ba1* and *cry39orf2* were present in the construct of CB/pHT1K-56Ba+39orf2 (Fig. 23, panel 1; Fig. 40A, panel 1). While there was no crystal observed in the construct of CB/pHT1K-56Ba (Fig. 23, Panel 2; Fig. 40A, Panel 2). Expression of Cry56Ba1 with Cry39ORF2 in construct of CB/pHT1K-56Ba+39orf2 was further confirmed by the SDS-PAGE (Fig. 37 lane7 and Fig. 40B, lane 1).

(ii) Over expression of the *cry56Ba1* operon stabilizes the Cry56Ba1 crystalline inclusions

To determine whether Cry56Ba1 crystals of uniform size and shape could result from over expression of the operon, the *cry56Ba1* (p1KSD-56Ba1; Fig. 23, Pane 4) or *cry56Ba1* and *cry*39orf2 (p1KSD-56Ba1+39orf2; Fig. 23, Pane 3) coding sequences were expressed using the strong chimeric *cyt1A-p*/STAB-SD expression system (Park *et al*., 1998) (Fig. 24 and 25). No inclusion were observed in CB/p1KSD-56Ba that lacked Cry39ORF2 (Fig. 40A, panel 4), whereas crystals of apparently spherical shape were visible in CB/p1KSD-56Ba+39orf2 (Fig. 40A, panel 3), and the size of inclusions in CB/p1KSD-56Ba+39orf2 was much bigger than the one in CB/pHT1K-56Ba+39orf2 (Fig. 40A, panel 1).

SDS-PAGE analysis showed that the amount of Cry56Ba1 synthesized by CB/p1KSD-56Ba+39orf2 (Fig. 40B, lanel 3) was much greater than that produced by CB/pHT1K-56Ba+39orf2 (Fig. 40B, lanel 1), while in recombinant CB/p1KSD-56Ba, which harbored the *cry56Ba1* gene alone, no protein of Cry56Ba1 was detected (Fig. 40B, lanel 4).

The interesting one is, there were big inclusions observed in CB/p1KSD-39orf2 (Fig. 40A, panel 5) while easily degraded and shown a more diffuse protein pattern in SDS-PAGE (Fig. 40B, lanel 5). The amino acid sequence alignment showed that 39ORF2 shared the highest level of identity with the C-terminal region of Cry4Aa (65%), Cry4Ba (65%), Cry7Ba1 (46%), Cry8Aa(46%), and Cry28Aa (46%) (Fig. 41).

Amino acid residues from 165 to 220 of 39ORF2 showed the least identity with C termini of other Cry proteins.

The evaluation of insecticidal activity of different constructs of Cry56Ba1 with Cry39ORF2 proteins against *C. pipiens molestus* and *C. pipiens pallens* 4th instar larvae showed that, the over-expression of recombinant CB/p1KSD-56Ba+39orf2 strain was approximately 14 fold more toxic than CB/pHT1K-56Ba+39orf2 (Table 16). The other 2 strains which contained *cry56Ba1* (CB/pHT1K-56Ba, CB/p1KSD-56Ba) showed no toxicity, even at 10 mg/ml. Interestingly, CB/p1KSD-39orf2, which produced inclusions but quickly degraded also showed no toxicity.

A

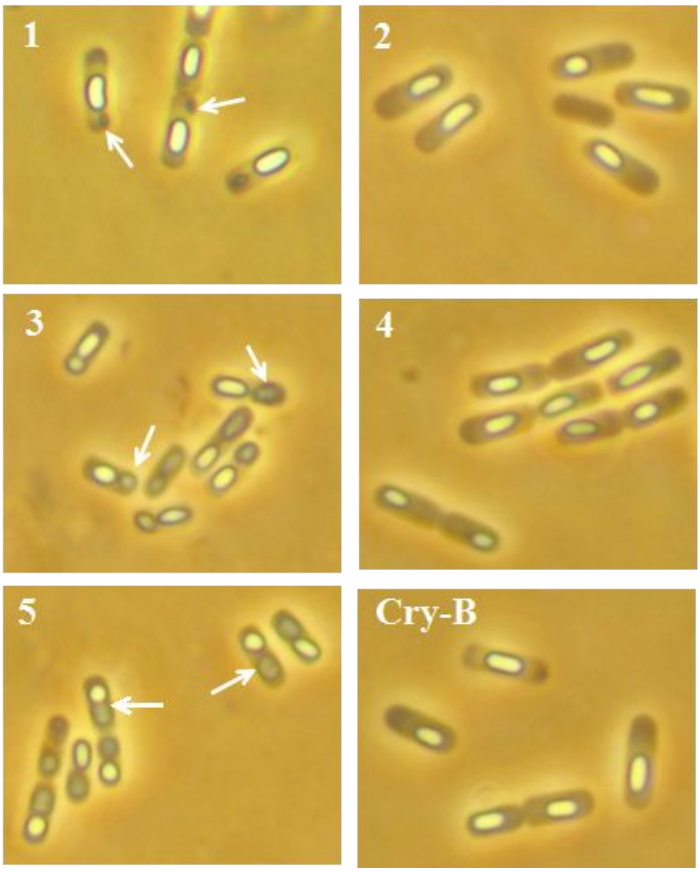


Fig. 40. Over-expression of *cry56Ba1* operon and functional analysis of 39ORF2.

(A) Phase contrast micrographs of the recombinant *B. thuringisnsis* Cry-B strains containing different constructs of *cry56Ba1* genes. 1, CB/pHT1K-56Ba+39orf2; 2, CB/pHT1K-56Ba; 3,CB/p1KSD-56Ba+39orf2; 4, CB/p1KSD-56Ba; 5, CB/p1KSD-39orf2; 6, wild type Cry$^-$B strain. Arrows indicate crystal.

B

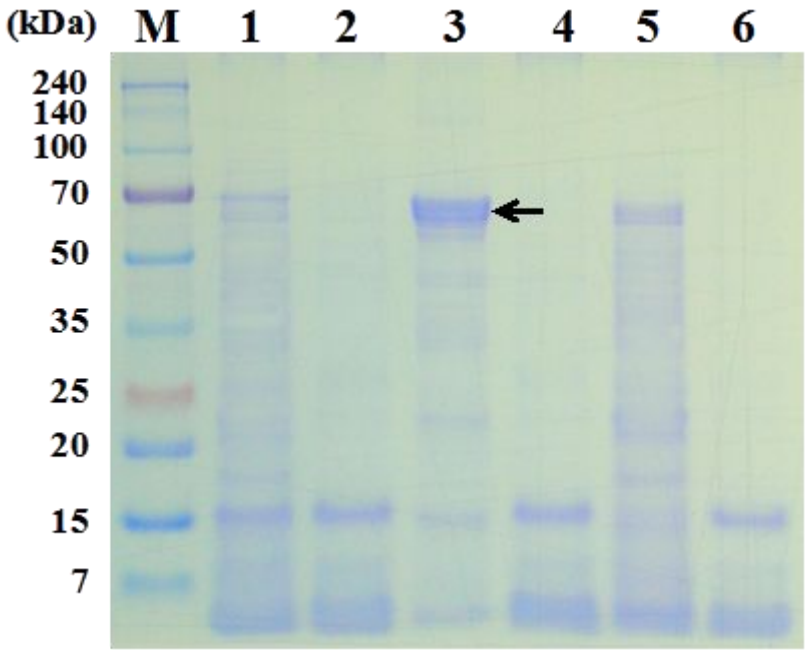


Fig. 40. Over-expression of *cry56Ba1* operon and functional analysis of 39ORF2.

(B) SDS–PAGE analysis of the recombinant *B. thuringiensis* Cry-B strains containing different *cry56Ba1* constructs. M, molecular mass marker; lane 1, CB/pHT1K-56Ba +39orf2; 2, CB/pHT1K-56Ba; 3,CB/p1KSD-56Ba+39orf2; 4, CB/p1KSD-56Ba; 5, CB/p1KSD- 39orf2; 6, wild type Cry$^-$B strain.

Table 16. Toxicity of recombinant strains of *B. thuringiensis* Cry-B against *Culex pipiens molestus* and *Cluex pipiens pallens* 4th instar larvae.

Strain	Toxin combination	*Culex pipiens molestus* LC_{50}[a] (mg/ml)	*Cluex pipiens pallens* LC_{50}[a] (mg/ml)
CB/pHT1K-56Ba+39orf2	Cry56Ba1 + Cry39orf2	1.02 (0.91 - 1.54)	1.26 (1.10 -1.42)
CB/pHT1K-56Ba	Cry56Ba1	Nt[b]	Nt[b]
CB/p1KSD-56Ba+39orf2	Cry56Ba1 + Cry39orf2	0.07 (0.05 -0.11)	0.08 (0.05 -0.12)
CB/p1KSD-56Ba1	Cry56Ba1	Nt[b]	Nt[b]
CB/p1KSD-39orf2	Cry39orf2	Nt[b]	Nt[b]
Cry-B	no	Nt[b]	Nt[b]

[a]LC_{50}: 50% lethal concentration (in μg) of freeze-dried spore–crystal complex per milliliter after 48 hours. The data are the total of three assays as determined by Probit analysis. [b]Nt, not toxic at 10 mg/ml; [c]FL_{95}: fiducial limits at P=0.95.

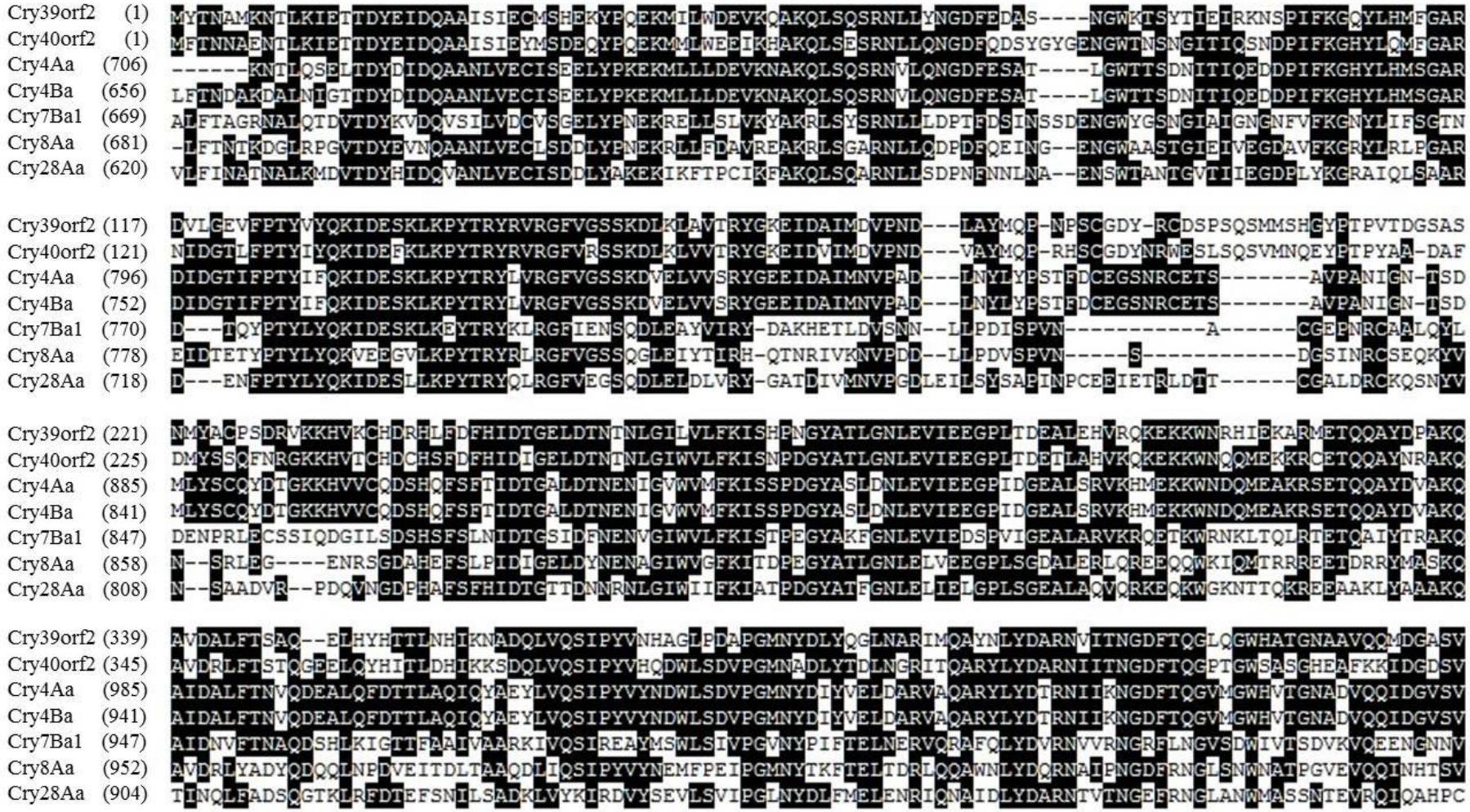

Cry39orf2 (1)
Cry40orf2 (1)
Cry4Aa (706)
Cry4Ba (656)
Cry7Ba1 (669)
Cry8Aa (681)
Cry28Aa (620)
Cry39orf2 (117)
Cry40orf2 (121)
Cry4Aa (796)
Cry4Ba (752)
Cry7Ba1 (770)
Cry8Aa (778)
Cry28Aa (718)
Cry39orf2 (221)
Cry40orf2 (225)
Cry4Aa (885)
Cry4Ba (841)
Cry7Ba1 (847)
Cry8Aa (858)
Cry28Aa (808)
Cry39orf2 (339)
Cry40orf2 (345)
Cry4Aa (985)
Cry4Ba (941)
Cry7Ba1 (947)
Cry8Aa (952)
Cry28Aa (904)

Fig. 41. Alignment of the orf2 amino acid sequence with C terminal regions of selected Cry proteins.

4. Discussion

The crystalline toxic proteins, δ-endotoxins, are predominantly synthesized as large, inactive protoxins that are activated by proteolysis in the insect gut (Gill *et al.*, 1992). The most common type Cry1 protein (about 130 kDa), consist of a N-terminal half containing the toxic portion of the molecule, released after ingestion by insect midgut proteases, and a C-terminal half important to crystallization (Baum and Malvar, 1995; Honee *et al.*, 1991). For example, Cry1, Cry4A, and Cry4B have molecular weights of 130 kDa to 140 kDa and are processed to active 65- to 70-kDa toxins (Gill *et al.*, 1992; Hofte and Whiteley, 1989). While Cry2A, Cry3A, Cry10A, and Cry11A are naturally truncated toxins and have molecular weights that range from 65 kDa to 80 kDa (Höfte and Whiteley, 1989). These proteins are correspond to the N-terminal half of the 130 kDa Cry type. Moreover, proteolytic cleavage at the N and C termini can also process these naturally truncated toxins to active 60 to 65 kDa toxins, as observed in Cry2A and Cry3A (Aronson, 1993).

Attempts to clone and express six *cry* genes (*cry19Bb1, cry20Bb1, cry73Aa* with *cry40orf2, cry27Ab1, cry4Aa* and *cry56Ba1* with *cry39orf2*) from *B.thuringiensis* subsp. *mogi*, four of them were failed to achieve sufficient levels of expression to allow the formation of a parasporal inclusion body as usually occurs with most *B.thuringiensis* endotoxins. Two of them, c*ry20Bb1* and *cry56Ba1 with cry39orf2*

were successfully formed inclusion body in the recombinant *B.thuringiensis* Cry$^-$B strains.

The deduced amino acid sequence of Cry20Bb1 shared a high level of identity (72.6%) with Cry20Aa from *B.thuringiensis* subsp. f*ukuokaensis* (Lee and Gill, 1997). To evaluate the mosquitocidal activity of the Cry20Bb1 protein, attempts were made to obtain purified parasporal inclusions. However, all attempts were unsuccessful since Cry20Bb1 rapidly degraded into tiny inclusions upon cell autolysis. These tiny inclusions contain mostly the smaller 50 and 30 kDa proteins, as determined by SDS-PAGE. Cry20Bb1 was degraded even when protease inhibitors were used during the isolation and purification procedures. The low mosquitocidal activity is not surprising since both the 50 and 30 kDa proteins have truncated domain I and III. Domain I forms pores in the insect midgut cell epithelium and is essential for insecticidal activity (Chen *et al*., 1995). Domain III is increasingly thought to play an essential role in insecticidal activity too (Chen *et al*., 1993).

Since wild-type *B. thuringiensis* subsp. *mogi* contains a number of proteins, each protein not only is a component of the inclusion but may also function to stabilize other proteins. This in part may explain the Cry20Bb1 stability in the wild-type strain but not in transformant Cry$^-$B strain.

Cry56Ba1 encodes a 659 amino acid protein containing only homology blocks 1–5. Homology blocks 6–8 are instead found in the protein Cry39ORF2, which is encoded

by the gene immediately following the intergenic region at the 3'-end of *cry56Ba1*. Eight similar *cry* gene pairs have been described previously in *B. thuringiensis*. The first genes of these eight pairs are *cry10Aa* (Thorne *et al.*, 1986), *cry19Aa* (Rosso and Delécuse, 1997), *cry24Ba* (Ohgushi *et al*., 2005), *cry30Ba* (Ito *et al*., 2006), *cry44Aa* (Ito *et al*., 2006), *cry5Ad* (Lenane *et al.*, 2008), *cry40Aa* and *cry40Ba* (GenBank accession numbers AB074414 and AB112346, respectively). Their upstream reading frames code for the Cry N-terminal domain, and the second frame found approximately 50-100 bp downstream codes for an apparent C-terminal domain that presumably has a function similar to that of ORF2 in protoxin aggregation and crystallization (Barboza-Corona *et al.,* 2012).

The primary genetic factors affecting insecticidal protein synthesis in *B. thuringiensis* are promoters, a 5'mRNA stabilizing sequence and 3' transcriptional termination sequences (Federici *et al*., 2010). The yield increases of Cry56Ba1 obtained with *cyt1AP*/STAB are likely due to higher gene expression resulting from the use of *cyt1A* promoter and especially to greater transcript stability conferred by the STAB-SD sequence (Agaisse and Lereclus, 1996). And the presence of 39orf2, also may stabilize the mRNA, or act as a chaperone to increase the stability of Cry56Ba1.

The toxicity of wild type *B. thuringiensis* subsp. *mogi* strain showed much higher than these recombinant strains. It could be explained by the possibility of synergistic action between different Cry proteins. Notwithstanding the complexity of form and

size, the protoxins are made as inactive protoxins and are activated by proteolysis to toxins. It is not known how these proteins of different size and amino acid sequence fold to generate common protease processing sites. There is virtually no information on the role of glycosylations in protease activation of toxin (Rukmini *et al.*, 2000). The process of activation appears to resemble that of mammalian gut proteases such as pepsinogen and trypsinogen in that a relatively small N-terminal peptide is removed. However, in case of protoxin activation, extensive C-terminal processing is involved and there are no internal cleavages generated within the toxic moiety during activation. It appears that conformational changes occurring during activation are rather subtle, affecting the tertiary structure but not the secondary structure of proteins (Choma and Kaplan, 1990; 1991). The polypeptide of toxic partiality in protoxin when compared to that of active toxin has different thermal unfolding properties.

Table S1. The CDS in pMOGI364 and their annotations.

CDS	Size (aa)	Strand	Annotation	Best hit in databases (GenBank no.)	(% aa identity)
1	473	+	hypothetical protein	hypothetical protein BCG9842_0219 [Bacillus cereus G9842] (YP_002454633.1)	99 in 473 aa
2	190	+	Ser/Thr protein phosphatase	metallophosphoesterase, calcineurin superfamily, putative [Bacillus cereus G9842] (YP_002454634.1)	99 in 190 aa
3	130	+	hypothetical protein	hypothetical protein bthur0007_54260 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111573.1)	95 in 130 aa
4	137	+	hypothetical protein	hypothetical protein IK9_05424 [Bacillus cereus VD166] (ZP_17621097.1)	85 in 136 aa
5	139	+	hypothetical protein	hypothetical protein BCG9842_0223 [Bacillus cereus G9842] (YP_002454637.1)	96 in 139 aa
6	146	+	hypothetical protein	hypothetical protein IK9_05422 [Bacillus cereus VD166] (ZP_17621095.1)	91 in 146 aa
7	147	-	hypothetical protein	hypothetical protein BCG9842_0225 [Bacillus cereus G9842] (YP_002454639.1)	96 in 139 aa
8	215	-	pseudogene		
9	139	+	hypothetical protein	hypothetical protein BCG9842_0227 [Bacillus cereus G9842] (YP_002454641.1)	99 in 139 aa
10	63	+	hypothetical protein	hypothetical protein BCG9842_0228 [Bacillus cereus G9842] (YP_002454642.1)	100 in 63 aa
11	96	+	hypothetical protein	hypothetical protein BCG9842_0229 [Bacillus cereus G9842] (YP_002454643.1)	99 in 96 aa
12	77	+	hypothetical protein	hypothetical protein BCG9842_0230 [Bacillus cereus G9842] (YP_002454644.1)	96 in 77 aa
13	141	+	hypothetical protein	hypothetical protein BCG9842_0231 [Bacillus cereus G9842] (YP_002454645.1)	96 in 141 aa
14	47	+	hypothetical protein	hypothetical protein BCG9842_0232 [Bacillus cereus G9842] (YP_002454646.1)	96 in 47 aa
15	158	+	hypothetical protein	hypothetical protein bthur0007_54170 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111564.1)	87 in 158 aa
16	241	-	S-layer protein	S-layer protein [Bacillus cereus G9842] (YP_002454648.1)	97 in 241 aa
17	288	+	ribosomal protein S1 domain protein	ribosomal protein S1 domain protein [Bacillus cereus G9842] (YP_002454649.1)	100 in 287 aa

18	321	+	hypothetical protein	hypothetical protein BCG9842_0236 [Bacillus cereus G9842] (YP_002454650.1)	100 in 321 aa
19	140	+	hypothetical protein	hypothetical protein BCG9842_0237 [Bacillus cereus G9842] (YP_002454651.1)	99 in 140 aa
20	595	+	hypothetical protein	hypothetical protein BCG9842_0238 [Bacillus cereus G9842] (YP_002454652.1)	97 in 595 aa
21	100	+	hypothetical protein	hypothetical protein BCG9842_0239 [Bacillus cereus G9842] (YP_002454653.1)	99 in 100 aa
22	121	+	hypothetical protein	hypothetical protein BCG9842_0240 [Bacillus cereus G9842] (YP_002454654.1)	98 in 120 aa
23	158	+	hypothetical protein	hypothetical protein BCG9842_0241 [Bacillus cereus G9842] (YP_002454655.1)	100 in 158 aa
24	135	+	single-strand binding protein	single-strand binding protein family [Bacillus cereus G9842] (YP_002454656.1)	96 in 135 aa
25	570	-	hypothetical protein	hypothetical protein BCG9842_0243 [Bacillus cereus G9842] (YP_002454657.1)	99 in 570 aa
26	484	+	hypothetical protein	hypothetical protein BCG9842_0244 [Bacillus cereus G9842] (YP_002454658.1)	99 in 484 aa
27	574	+	RpiR family transcriptional regulator	hypothetical protein BCG9842_0245 [Bacillus cereus G9842] (YP_002454659.1)	99 in 574 aa
28	129	-	hypothetical protein	hypothetical protein BCG9842_0246 [Bacillus cereus G9842] (YP_002454660.1)	99 in 129 aa
29	154	-	hypothetical protein	hypothetical protein BCG9842_0247 [Bacillus cereus G9842] (YP_002454661.1)	98 in 153 aa
30	248	+	hypothetical protein	hypothetical protein BCG9842_0248 [Bacillus cereus G9842] (YP_002454662.1)	97 in 248 aa
31	199	+	hypothetical protein	hypothetical protein BCG9842_0249 [Bacillus cereus G9842] (YP_002454663.1)	98 in 199 aa
32	361	+	hypothetical protein	hypothetical protein BCG9842_0250 [Bacillus cereus G9842] (YP_002454664.1)	99 in 361 aa
33	173	-	hypothetical protein	hypothetical protein BCG9842_0251 [Bacillus cereus G9842] (YP_002454665.1)	98 in 166 aa
34	51	+	hypothetical protein	hypothetical protein BCG9842_0252 [Bacillus cereus G9842] (YP_002454666.1)	100 in51 aa
35	75	+	hypothetical protein	hypothetical protein bthur0001_57990 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149204.1)	69 in 55 aa
36	516	+	hypothetical protein	hypothetical protein BCG9842_0254 [Bacillus cereus G9842] (YP_002454668.1)	89 in 437 aa
37	188	+	hypothetical protein	hypothetical protein bthur0007_59340 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112055.1)	96 in 187 aa
38	1116	+	DEAD/DEAH box helicase DEAD-like helicase	helicase conserved C- domain protein [Bacillus cereus G9842] (YP_002454670.1)	99 in 1116 aa

39	287	+	hypothetical protein	hypothetical protein BCG9842_0005 [Bacillus cereus G9842] (YP_002454671.1)	100 in 287 aa
40	332	+	acetyltransferase	YdjC [Bacillus cereus G9842] (YP_002454672.1)	99 in 332 aa
41	450	+	Sporulation kinase	multi-sensor signal transduction histidine kinase, putative [Bacillus cereus G9842] (YP_002454673.1)	99 in 450 aa
42	191	+	NarL family DNA-binding response regulator	Two-component response regulator YhcZ, putative [Bacillus cereus G9842] (YP_002454674.1)	99 in 191 aa
43	60	+	HTH-type transcriptional regulator SinR	HTH-type transcriptional regulator SinR [Bacillus cereus G9842] (YP_002454675.1)	98 in 60 aa
44	391	+	hypothetical protein	hypothetical protein BCG9842_0010 [Bacillus cereus G9842] (YP_002454676.1)	98 in 391 aa
45	50	+	hypothetical protein	hypothetical protein BCG9842_0011 [Bacillus cereus G9842] (YP_002454677.1)	96 in 50 aa
46	171	+	hypothetical protein	hypothetical protein BCG9842_0012 [Bacillus cereus G9842] (YP_002454678.1)	99 in 169 aa
47	64	+	hypothetical protein	hypothetical protein BCG9842_0012 [Bacillus cereus G9842] (YP_002454678.1)	98 in 64 aa
48	179	+	hypothetical protein	hypothetical protein bthur0001_54190 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148849.1)	98 in 197 aa
49	191	+	hypothetical protein	hypothetical protein BCG9842_0014 [Bacillus cereus G9842] (YP_002454680.1)	97 in 191 aa
50	548	+	DNA polymerase III subunit gamma/tau	DNA polymerase III subunit gamma/tau [Bacillus cereus G9842] (YP_002454681.1)	99 in 548 aa
51	55	-	hypothetical protein	hypothetical protein bthur0007_59220 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112043.1)	100 in 55 aa
52	121	-	hypothetical protein	YtvB [Bacillus cereus G9842] (YP_002454683.1)	99 in 121 aa
53	152	+	hypothetical protein	hypothetical protein bthur0007_59200 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112041.1)	97 in 152 aa
54	199	+	hypothetical protein	hypothetical protein IK9_05618 [Bacillus cereus VD166] (ZP_17621291.1)	96 in 199 aa
55	51	+	hypothetical protein	hypothetical protein bthur0007_59180 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112039.1)	100 in 51 aa
56	663	+	phosphoadenosine phosphosulfate reductase	hypothetical protein BCG9842_0022 [Bacillus cereus G9842] (YP_002454688.1)	98 in 388 aa

57	388	+	sulfurtransferase DndC	hypothetical protein BCG9842_0022 [Bacillus cereus G9842] (YP_002454688.1)	98 in 388 aa
58	174	-	signal peptidase I	signal peptidase I [Bacillus cereus G9842] (YP_002454689.1)	93 in 174 aa
59	163	+	oligopeptide ABC transporter ATP-binding protein	hypothetical protein bthur0007_59140 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112035.1)	96 in 162 aa
60	122	+	cell division protein SepF	hypothetical protein BCG9842_0025 [Bacillus cereus G9842] (YP_002454691.1)	95 in 122 aa
61	624	+	HNH endonuclease	hypothetical protein BCG9842_0026 [Bacillus cereus G9842] (YP_002454692.1)	81 in 344 aa
62	219	+	group-specific protein	hypothetical protein IK9_05610 [Bacillus cereus VD166] (ZP_17621283.1)	94 in 224 aa
63	415	+	glutathionylspermidine synthase	Glutathionylspermidine synthase [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148930.1)	97 in 415 aa
64	75	-	hypothetical protein	hypothetical protein IK9_05608 [Bacillus cereus VD166] (ZP_17621281.1)	93 in 45 aa
65	157	+	hypothetical protein	hypothetical protein BCG9842_0029 [Bacillus cereus G9842] (YP_002454695.1)	99 in 157 aa
66	214	+	thermonuclease family protein	thermonuclease family protein [Bacillus cereus G9842] (YP_002454696.1)	99 in 214 aa
67	89	+	actin binding protein	hypothetical protein IK9_05605 [Bacillus cereus VD166] (ZP_17621278.1)	100 in 89 aa
68	201	+	hypothetical protein	hypothetical protein BCG9842_0032 [Bacillus cereus G9842] (YP_002454698.1)	100 in 200 aa
69	117	+	hypothetical protein	hypothetical protein BCG9842_0033 [Bacillus cereus G9842] (YP_002454699.1)	98 in 117 aa
70	214	-	signal peptidase I	signal peptidase I [Bacillus cereus VD166] (ZP_17621275.1)	99 in 214 aa
71	286	+	RNA polymerase sigma factor RpoD	RNA polymerase sigma factor, sigma-70 family [Bacillus cereus G9842] (YP_002454701.1)	97 in 284 aa
72	127	+	hypothetical protein	hypothetical protein bthur0001_55210 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148937.1)	94 in 127 aa
73	84	+	recombination protein RecR	hypothetical protein IK9_05599 [Bacillus cereus VD166] (ZP_17621272.1)	92 in 84 aa
74	152	+	hypothetical protein	hypothetical protein bthur0001_55230 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148939.1)	94 in 152 aa
75	43	+	hypothetical protein	hypothetical protein BCG9842_0039 [Bacillus cereus G9842] (YP_002454705.1)	98 in 43 aa

76	81	+	hypothetical protein	hypothetical protein BCG9842_0039 [Bacillus cereus G9842] (YP_002454705.1)	99 in 81 aa
77	90	+	hypothetical protein	hypothetical protein bthur0007_58950 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112016.1)	98 in 90 aa
78	313	+	hypothetical protein	hypothetical protein bthur0007_58940 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112015.1)	98 in 313 aa
79	583	+	hypothetical protein	hypothetical protein BCG9842_0042 [Bacillus cereus G9842] (YP_002454708.1)	99 in 583 aa
80	323	+	DNA polymerase III subunit delta	DNA polymerase III, delta subunit [Bacillus cereus G9842] (YP_002454709.1)	100 in 232 aa
81	179	+	hypothetical protein	hypothetical protein BCG9842_0044 [Bacillus cereus G9842] (YP_002454710.1)	99 in 179 aa
82	57	+	pseudogene		
83	250	-	transposase IstB	hypothetical protein IK5_00386 [Bacillus cereus VD154] (ZP_17603283.1)	98 in 250 aa
84	431	-	transposase IstA	transposase for insertion sequence element IS232 [Bacillus cereus VD154] (ZP_17609012.1)	91 in 431 aa
85	188	+	pseudogene		
86	83	+	host factor Hfq	host factor Hfq [Bacillus cereus G9842] (YP_002454712.1)	100 in 83 aa
87	133	+	hypothetical protein	hypothetical protein BCG9842_0047 [Bacillus cereus G9842] (YP_002454713.1)	99 in 133 aa
88	214	+	hypothetical protein	hypothetical protein BCG9842_0048 [Bacillus cereus G9842] (YP_002454714.1)	98 in 214 aa
89	152	+	hypothetical protein	hypothetical protein BCG9842_0049 [Bacillus cereus G9842] (YP_002454715.1)	98 in 152 aa
90	37	+	hypothetical protein	hypothetical protein BCG9842_0050 [Bacillus cereus G9842] (YP_002454716.1)	100 in 37 aa
91	96	+	hypothetical protein	hypothetical protein BCG9842_0051 [Bacillus cereus G9842] (YP_002454717.1)	96 in 96 aa
92	92	+	transcriptional repressor PagR	transcriptional regulator, ArsR family [Bacillus cereus G9842] (YP_002454718.1)	100 in 92 aa
93	160	+	hypothetical protein	hypothetical protein BCG9842_0053 [Bacillus cereus G9842] (YP_002454719.1)	99 in 160 aa
94	124	+	hypothetical protein	hypothetical protein BCG9842_0054 [Bacillus cereus G9842] (YP_002454720.1)	98 in 124 aa
95	97	+	hypothetical protein	hypothetical protein BCG9842_0055 [Bacillus cereus G9842] (YP_002454721.1)	100 in 97 aa
96	79	+	hypothetical protein	hypothetical protein IK9_05579 [Bacillus cereus VD166] (ZP_17621252.1)	88 in77 aa

97	76	+	hypothetical protein	hypothetical protein IK9_05578 [Bacillus cereus VD166] (ZP_17621251.1)	89 in 76 aa
98	380	-	integrase-recombinase	hypothetical protein IK9_05577 [Bacillus cereus VD166] (ZP_17621250.1)	94 in 380 aa
99	276	+	protein translocase subunit secA	protein translocase subunit secA [Bacillus cereus VD166] (ZP_17621249.1)	91 in 276 aa
100	101	+	hypothetical protein	hypothetical protein IK9_05575 [Bacillus cereus VD166] (ZP_17621248.1)	90 in 101 aa
101	92	+	transition state regulatory protein AbrB	Transition state regulatory protein AbrB [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111996.1)	97 in 92.aa
102	72	+	membrane protein	membrane protein [Planococcus antarcticus DSM 14505] (ZP_10206699.1)	70 in 71 aa
103	238	+	ABC transporter ATP-binding protein	ABC transporter, ATP-binding protein [Planococcus antarcticus DSM 14505] (ZP_10206700.1)	74 in 238 aa
104	457	+	membrane protein	hypothetical protein A1A1_01488 [Planococcus antarcticus DSM 14505] (ZP_10206701.1)	64 in 458 aa
105	91	+	hypothetical protein	hypothetical protein A1A1_01493 [Planococcus antarcticus DSM 14505] (ZP_10206702.1)	69 in 85 aa
106	125	+	membrane protein	membrane protein [Planococcus antarcticus DSM 14505] (ZP_10206703.1)	54 in 125 aa
107	145	+	disulfide formation protein C	hypothetical protein IE5_05407 [Bacillus cereus BAG3X2-2] (ZP_17404749.1)	67 in 141 aa
108	324	+	DNA-methyltransferase	DNA-methyltransferase [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148963.1)	90 in 330 aa
109	92	+	membrane protein	hypothetical protein IK9_05571 [Bacillus cereus VD166] (ZP_17621244.1)	93 in 92 aa
110	228	+	hypothetical protein	hypothetical protein ICU_04726 [Bacillus cereus BAG2X1-1] (ZP_17376233.1)	93 in 228 aa
111	71	+	Small, acid-soluble spore protein C	Small, acid-soluble spore protein C [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111991.1)	97 in 71 aa
112	140	+	hypothetical protein	hypothetical protein BCG9842_0066 [Bacillus cereus G9842] (YP_002454732.1)	89 in 140 aa
113	62	+	hypothetical protein	hypothetical protein bthur0007_58680 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111989.1)	80 in 61 aa

114	190	+	hypothetical protein	conserved hypothetical protein [Bacillus cereus AH1134] (ZP_03233313.1)	80 in 109 aa
115	73	+	hypothetical protein	hypothetical protein II7_05369 [Bacillus cereus MSX-A12] (ZP_17548393.1)	95 in 73 aa
116	55	+	hypothetical protein	hypothetical protein IG1_05794 [Bacillus cereus HD73] (ZP_17484757.1)	78 in 55 aa
117	102	+	hypothetical protein	hypothetical protein bthur0008_37170 [Bacillus thuringiensis serovar berliner ATCC 10792] (ZP_04103634.1)	91 in 101 aa
118	241	+	hypothetical protein	hypothetical protein IC1_05843 [Bacillus cereus VD022] (ZP_17341366.1)	93 in 241 aa
119	72	+	hypothetical protein	hypothetical protein IIE_06160 [Bacillus cereus VD045] (ZP_17566835.1)	91 in 66 aa
120	65	+	hypothetical protein	hypothetical protein bthur0007_58660 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111987.1)	85 in 54 aa
121	150	+	vacuolar protein-sorting -associated protein 36	hypothetical protein bthur0007_58650 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111986.1)	88 in 150 aa
122	326	+	MerR family transcriptional regulator	transcriptional regulator, MerR [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111984.1)	97 in 326 aa
123	306	+	WXG100 family type VII secretion target	hypothetical protein bthur0007_58620 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111983.1)	99 in 306 aa
124	119	+	hypothetical protein	hypothetical protein bthur0007_58610 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1]	100 in 119 aa
125	90	-	DNA-binding protein HU	DNA-binding protein HU [Bacillus cereus VD166] (ZP_17621231.1)	96 in 90 aa
126	81	+	hypothetical protein	hypothetical protein bthur0007_61170 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112230.1)	98 in 81 aa
127	52	-	hypothetical protein	hypothetical protein bthur0007_55320 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111679.1)	82 in 50 aa
128	357	-	response regulator aspartate phosphatase	hypothetical protein bthur0001_55580 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148974.1)	100 in 357 aa

129	147	+	ArsR family transcriptional regulator	hypothetical protein bthur0001_55590 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148975.1)	99 in 147 aa
130	64	+	hypothetical protein	hypothetical protein bthur0001_55600 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148976.1)	100 in 64 aa
131	327	+	hypothetical protein	hypothetical protein bthur0001_55610 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148977.1)	99 in 327 aa
132	133	+	hypothetical protein	hypothetical protein bthur0007_61210 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112234.1)	98 in 133 aa
133	244	+	hypothetical protein	hypothetical protein bthur0001_55630 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148979.1)	99 in 241 aa
134	82	+	hypothetical protein	hypothetical protein IK9_05548 [Bacillus cereus VD166] (ZP_17621221.1)	100 in 82 aa
135	60	-	hypothetical protein	hypothetical protein BCG9842_0085 [Bacillus cereus G9842] (YP_002454751.1)	100 in 60 aa
136	185	+	histidinol-phosphate phosphatase domain-containing protein	histidinol-phosphate phosphatase domain-containing protein [Bacillus cereus VD166] (ZP_17621219.1)	95 in 185 aa
137	141	+	thioredoxin	hypothetical protein BCG9842_0087 [Bacillus cereus G9842] (YP_002454753.1)	97 in 141 aa
138	173	+	hypothetical protein	hypothetical protein bthur0001_55670 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148983.1)	98 in 173 aa
139	152	+	hypothetical protein	hypothetical protein BCG9842_0089 [Bacillus cereus G9842] (YP_002454755.1)	98 in 152 aa
140	157	+	hypothetical protein	hypothetical protein BCG9842_0090 [Bacillus cereus G9842] (YP_002454756.1)	99 in 157 aa
141	98	+	hypothetical protein	hypothetical protein BCG9842_0091 [Bacillus cereus G9842] (YP_002454757.1)	98 in 98 aa
142	273	+	UvrD/REP helicase	helicase, UvrD/Rep family [Bacillus cereus G9842] (YP_002454758.1)	99 in 270 aa
143	612	+	group II intron-encoded protein LtrA	group II intron-encoded protein LtrA [Bacillus cereus 03BB108] (ZP_03114863.1)	89 in 612 aa
144	377	+	ATP-dependent DNA helicase PcrA	ATP-dependent DNA helicase PcrA [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112242.1)	96 in 377 aa

145	326	+	DNA polymerase III subunit delta'	DNA polymerase III subunit delta' [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112243.1)	98 in 326 aa
146	87	+	hypothetical protein	hypothetical protein BCG9842_0094 [Bacillus cereus G9842] (YP_002454760.1)	61 in 94 aa
147	312	+	RimK domain-containing protein ATP-grasp	hypothetical protein SchaN1_13993 [Streptomyces chartreusis NRRL 12338] (ZP_09954292.1)	31 in 314 aa
148	327	-	intracellular serine protease	Intracellular serine protease [Bacillus cereus 172560W] (ZP_04309338.1)	65 in 327 aa
149	305	-	carboxypeptidase domain-containing protein	hypothetical protein PpisJ2_04638 [pseudogenealteromonas piscicida JCM 20779] (ZP_10288245.1)	31 in 275 aa
150	75	+	Rev-Erb beta 2	hypothetical protein bthur0001_56290 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149037.1)	99 in 75 aa
151	151	+	hypothetical protein	hypothetical protein bthur0001_56280 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149036.1)	98 in 151 aa
152	121	+	hypothetical protein	hypothetical protein bthur0001_56270 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149035.1)	97 in 121 aa
153	581	+	hypothetical protein	hypothetical protein bthur0007_61360 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112249.1)	94 in 581 aa
154	81	+	hypothetical protein	hypothetical protein IK9_05532 [Bacillus cereus VD166] (ZP_17621205.1)	94 in 81 aa
155	438	+	DNA (cytosine-5-) -methyltransferase	DNA (cytosine-5-)-methyltransferase [Bacillus cereus VD166] (ZP_17621204.1)	99 in 438 aa
156	209	+	thermonuclease	thermonuclease [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149031.1)	99 in 209 aa
157	299	+	foldase protein PrsA	foldase protein PrsA [Bacillus cereus G9842] (YP_002454772.1)	98 in 299 aa
158	114	+	hypothetical protein	hypothetical protein bthur0007_61430 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112256.1)	94 in 114 aa

159	58	+	hypothetical protein	hypothetical protein bthur0007_61440 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112257.1)	100 in 58 aa
160	156	+	hypothetical protein	hypothetical protein bthur0007_61450 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112258.1)	96 in 156 aa
161	131	+	hypothetical protein	hypothetical protein bthur0007_61460 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112259.1)	92 in 131 aa
162	87	+	hypothetical protein	hypothetical protein bthur0007_61470 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112260.1)	97 in 87 aa
163	132	+	hypothetical protein	hypothetical protein bthur0007_61490 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112262.1)	87 in 132 aa
164	243	+	hypothetical protein	hypothetical protein IE1_05529 [Bacillus cereus BAG3O-2] (ZP_17393345.1)	79 in 243 aa
165	135	+	hypothetical protein	hypothetical protein IK9_05524 [Bacillus cereus VD166] (ZP_17621197.1)	89 in 135 aa
166	109	+	hypothetical protein	hypothetical protein IK9_05522 [Bacillus cereus VD166] (ZP_17621195.1)	95 in 109 aa
167	167	-	hypothetical protein	hypothetical protein bthur0007_61520 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112265.1)	99 in 167 aa
168	463	-	Type II DNA-methyltransferase	Type II DNA-methyltransferase [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112266.1)	98 in 463 aa
169	116	+	hypothetical protein	hypothetical protein bthur0001_56920 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149100.1)	95 in 116 aa
170	144	+	hypothetical protein	hypothetical protein bthur0001_56910 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149099.1)	93 in 144 aa
171	119	+	hypothetical protein	hypothetical protein BCG9842_0121 [Bacillus cereus G9842] (YP_002454787.1)	93 in 119
172	337	-	primosomal protein DnaI	primosomal protein DnaI [Bacillus cereus G9842] (YP_002454788.1)	98 in 337 aa
173	86	-	hypothetical protein	hypothetical protein BCG9842_0123 [Bacillus cereus G9842] (YP_002454789.1)	99 in 86 aa

174	65	-	hypothetical protein	hypothetical protein IK9_05515 [Bacillus cereus VD166] (ZP_17621188.1)	94 in 65 aa
175	79	+	hypothetical protein	hypothetical protein bthur0001_56860 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (WP_000043819.1)	94 in 79 aa
176	62	+	RNA chaperone Hfq	Host factor-I protein [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149093.1)	98 in 62 aa
177	459	+	hypothetical protein	hypothetical protein bthur0007_61590 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112272.1)	76 in 459 aa
178	99	+	ArsR family transcriptional regulator	transcriptional regulator, ArsR family [Bacillus cereus G9842] (YP_002454793.1)	99 in 99 aa
179	133	+	hypothetical protein	hypothetical protein bthur0001_56820 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149090.1)	99 in 133 aa
180	72	+	hypothetical protein	ypothetical protein IK9_05507 [Bacillus cereus VD166] (ZP_17621180.1)	100 in 72 aa
181	96	+	hypothetical protein	hypothetical protein BCG9842_0135 [Bacillus cereus G9842] (YP_002454801.1)	94 in 96 aa
182	361	+	integrase/recombinase, phage integrase family protein	integrase/recombinase, phage integrase family protein [Bacillus cereus G9842] (YP_002454802.1)	99 in 361 aa
183	174	-	hypothetical protein	hypothetical protein bthur0001_56800 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149088.1)	98 in 174 aa
184	239	-	hypothetical protein	hypothetical protein BCG9842_0138 [Bacillus cereus G9842] (YP_002454804.1)	96 in 239 aa
185	115	-	ribosome biogenesis GTPase rsgA	hypothetical protein BCG9842_0139 [Bacillus cereus G9842] (YP_002454805.1)	96 in 114 aa
186	420	-	DNA-damage repair protein	DNA-damage repair protein [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04111654.1)	92 in 421 aa
187	394	+	RES domain-containing protein	hypothetical protein IK9_05501 [Bacillus cereus VD166] (ZP_17621174.1)	96 in 392 aa
188	118	-	hypothetical protein	TrsE, putative [Bacillus cereus G9842] (YP_002454807.1)	56 in 118 aa
189	60	-	hypothetical protein	hypothetical protein BCG9842_0142 [Bacillus cereus G9842] (YP_002454808.1)	97 in 60 aa
190	167	-	hypothetical protein	hypothetical protein BCG9842_0143 [Bacillus cereus G9842] (YP_002454809.1)	81 in 167 aa

191	131	-	hypothetical protein	hypothetical protein BCG9842_0144 [Bacillus cereus G9842] (YP_002454810.1)	94 in 131 aa
192	73	-	hypothetical protein	hypothetical protein bthur0001_56730 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149081.1)	96 in 73 aa
193	137	-	hypothetical protein	hypothetical protein bthur0013_54750 [Bacillus thuringiensis IBL 200] (ZP_04075141.1)	71 in 128 aa
194	173	-	hypothetical protein	hypothetical protein IGK_05536 [Bacillus cereus HuB4-10] (ZP_17519835.1)	69 in 165 aa
195	61	-	hypothetical protein	hypothetical protein BCG9842_0154 [Bacillus cereus G9842] (YP_002454820.1)	83 in 60 aa
196	182	-	hypothetical protein	hypothetical protein IK9_05488 [Bacillus cereus VD166] (ZP_17621161.1)	88 in 180 aa
197	465	-	hypothetical protein	hypothetical protein BCG9842_0156 [Bacillus cereus G9842] (YP_002454822.1)	99in 465 aa
198	96	-	hypothetical protein	hypothetical protein BCG9842_0157 [Bacillus cereus G9842] (YP_002454823.1)	98 in 96 aa
199	330	+	hypothetical protein	hypothetical protein bthur0007_54930 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111640.1)	97 in 330
200	59	-	hypothetical protein	hypothetical protein bthur0007_54920 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111639.1)	98 in 59 aa
201	478	+	transposase for insertion sequence element IS231B	transposase for insertion sequence element IS231B [Bacillus thuringiensis MC28] (YP_006815473.1)	95 in 478 aa
202	57	-	hypothetical protein	hypothetical protein BCG9842_0161 [Bacillus cereus G9842] (YP_002454827.1)	100 in 57 aa
203	293	-	DNA integration/recombination/ invertion protein	DNA integration/recombination/invertion protein [Bacillus cereus G9842] (YP_002454828.1)	98 in 293 aa
204	136	-	hypothetical protein	hypothetical protein bthur0001_56580 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149066.1)	99 in 136 aa
205	163	-	hypothetical protein	hypothetical protein BCG9842_0164 [Bacillus cereus G9842] (YP_002454830.1)	95 in 163 aa
206	245	-	hypothetical protein	hypothetical protein BCG9842_0165 [Bacillus cereus G9842] (YP_002454831.1)	90 in 245 aa
207	313	-	pseudogene	(DNA topoisomerase III, 146kbp fragment insertion site)	

208	106	-	pseudogene		
209	324	-	transmembrane anti-sigma factor	hypothetical protein BMQ_pBM50024 [Bacillus megaterium QM B1551] (YP_003566043.1)	73 in 325 aa
210	176	-	RNA polymerase sigma factor	RNA polymerase sigma factor [Bacillus cereus Rock3-28] (ZP_04236331.1)	84 in 175 aa
211	199	-	ABC transporter permease protein	ABC transporter permease protein [Bacillus thuringiensis IBL 200] (ZP_04075478.1)	74 in 198 aa
212	621	-	ABC transporter permease protein	hypothetical protein IE9_05147 [Bacillus cereus BAG4X12-1] (ZP_17415947.1)	71 in 621 aa
213	317	-	ABC transporter ATP-binding protein	hypothetical protein ICE_05420 [Bacillus cereus BAG1X1-2] (ZP_17364930.1)	72 in 227
214	100	-	hypothetical protein	hypothetical protein IC1_06304 [Bacillus cereus VD022] (ZP_17341827.1)	49 in 97 aa
215	111	-	hypothetical protein	hypothetical protein bthur0011_5800 [Bacillus thuringiensis serovar huazhongensis BGSC 4BD1] (ZP_04082919.1)	44 in 107 aa
216	480	-	Calcineurin-like phosphoesterase	hemagglutinin-related protein [Bacillus thuringiensis serovar israelensis] (YP_001573868.1)	44 in 490 aa
217	85	-	hypothetical protein	hypothetical protein MC28_D172 [Bacillus thuringiensis MC28] (YP_006815460.1)	84 in 85 aa
218	98	-	pseudogene		
219	194	+	hypothetical protein	hypothetical protein bthur0003_63770 [Bacillus thuringiensis serovar thuringiensis str. T01001] (ZP_04137140.1)	81 in 194 aa
220	101	+	ArsR family transcriptional regulator	hypothetical protein IKM_06008 [Bacillus cereus VDM022] (ZP_17641027.1)	85 in 101 aa
221	93	+	AbrB family transcriptional regulator	AbrB family transcriptional regulator [Bacillus cereus VD045] (ZP_17566700.1)	88 in 93 aa
222	62	+	RNA chaperone Hfq	HfQ protein (RNA-binding protein) [Bacillus cereus Q1] (YP_002533362.1)	90 in 62 aa
223	65	+	hypothetical protein	hypothetical protein YBT020_27704 [Bacillus thuringiensis serovar finitimus YBT-020] (YP_005569122.1)	91 in 65 aa
224	197	-	camelysin	camelysin [Bacillus thuringiensis serovar thuringiensis str. T01001] (ZP_04137145.1)	96 in 197 aa

225	112	-	hypothetical protein	hypothetical protein bthur0007_9040 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04107101.1)	89 in 101 aa
226	109	-	hypothetical protein	hypothetical protein IKM_06002 [Bacillus cereus VDM022] (ZP_17641021.1)	39 in 99 aa
227	241	-	hypothetical protein	hypothetical protein BFZC1_00135 [Lysinibacillus fusiformis ZC1] (ZP_07047732.1)	41 in 189 aa
228	549	-	Phage integrase family protein	Phage integrase family protein [Lysinibacillus fusiformis ZC1] (ZP_07047733.1)	33 in 564 aa
229	231	-	cell envelope-bound metalloprotease	hypothetical protein bcere0013_32640 [Bacillus cereus BDRD-ST26] (ZP_04268721.1)	28 in 428 aa
230	413	-	phage integrase family site-specific recombinase	hypothetical protein IIK_01395 [Bacillus cereus VD102] (ZP_17580707.1)	32 in 640 aa
231	779	-	peptidase S8 and S53 subtilisin kexin sedolisin	hypothetical protein ACD_77C00511G0002 [uncultured bacterium] (EKD30588.1)	48 in 316 aa
232	383	-	ATPase AAA	ATPase [Clostridium botulinum A str. ATCC 3502] (YP_001254814.1)	94 in 673 aa
233	673	-	DNA topoisomerase III	DNA topoisomerase III [Bacillus cereus HuA2-1] (ZP_17486349.1)	78 in 76 aa
234	236	-	pseudogene		
235	54	-	pseudogene		
236	261	+	pseudogene		
237	131	-	HTH domain-containing DNA-binding protein	hypothetical protein bthur0006_5990 [Bacillus thuringiensis serovar kurstaki str. T03a001] (ZP_04113288.1)	87 in 105 aa
238	684	+	cry19Bb1	Pesticidal crystal protein cry19Ba (O86170.1)	74 in 690 aa
239	172	+	IS3-family transposase OrfB	transposase, orfA ISRSO11-related [Bacillus cereus E33L] (YP_245796.1)	100 in 172 aa
240	265	+	ISPsy9 transposase OrfA	transposase orfB, IS150-related protein [Bacillus thuringiensis HD-771] (YP_006602580.1)	100 in 265 aa
241	168	-	pseudogene		

242	120	-	pseudogene		
243	96	-	pseudogene		
244	306	-	pseudogene		
245	79	-	Phage protein	hypothetical protein MC28_1517 [Bacillus thuringiensis MC28] (YP_006828338.1)	59 in 104 aa
246	71	-	Phage protein		
247	324	-	pseudogene		
248	50	-	Phage protein	hypothetical protein II3_05736 [Bacillus cereus MC67] (ZP_17536834.1)	66 in 50 aa
249	351	+	pseudogene		
250	65	+	hypothetical protein	hypothetical protein IIO_04055 [Bacillus cereus VD115] (ZP_17594563.1)	89 in 63 aa
251	80	+	pseudogene		
252	269	-	IS3-family transposase OrfB	hypothetical protein IKA_00265 [Bacillus cereus VD169] (ZP_17622048.1)	96 in 269 aa
253	176	-	ISPsy9 transposase OrfA	hypothetical protein II3_00131 [Bacillus cereus MC67] (ZP_17531229.1)	97 in 172 aa
254	105	-	pseudogene		
255	259	-	pseudogene		
256	89	-	pseudogene		
257	26	-	pseudogene		
258	602	-	RNA-directed DNA polymerase	RNA-directed DNA polymerase [Enterococcus faecalis CH188] (ZP_05585309.1)	57 in 602
259	59	-	Holin	Holin [Bacillus thuringiensis serovar kurstaki str. T03a001] (ZP_04118140.1)	60 in 88 aa
260	74	-	Cof-like hydrolase	hypothetical protein II5_02389 [Bacillus cereus MSX-A1] (ZP_17539261.1)	96 in 74 aa
261	497	-	Cry40orf2	Cry40-like protein [Bacillus thuringiensis MC28] (YP_006815592.1）	74 in 498 aa
262	669	-	Cry73Aa	crystal protein [Bacillus thuringiensis serovar vazensis] (AFM37573.1)	86 in 669 aa
263	159	+	pseudogene		
264	334	+	pseudogene		

265	45	-	hypothetical protein	hypothetical protein bthur0014_3770 [Bacillus thuringiensis IBL 4222] (ZP_04063421.1)	72 in 46 aa
266	45	-	hypothetical protein	hypothetical protein [Bacillus anthracis str. H9401] (YP_006207480.1)	80 in 45 aa
267	100	-	pseudogene		
268	107	+	ISPsy9, transposase OrfA	hypothetical protein IKA_05207 [Bacillus cereus VD169] (ZP_17626990.1)	100 in 100 aa
269	279	+	ISPsy9, transposase OrfB	hypothetical protein IKA_05206 [Bacillus cereus VD169] (ZP_17626989.1)	99 in 279 aa
270	54	-	Phage protein	Phage protein [Bacillus sp. GeD10] (CCW09303.1)	80 in 54 aa
271	71	-	Phage protein	Phage protein [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00742080.1)	73 in 71 aa
272	138	-	hypothetical protein	Hypothetical cytosolic protein [Bacillus thuringiensis IBL 4222] (ZP_04063412.1)	95 in 138 aa
273	133	-	pseudogene		
274	171	-	pseudogene		
275	41	-	pseudogene		
276	118	-	pseudogene		
277	62	-	pseudogene		
278	104	-	hypothetical protein	hypothetical protein IGO_05580 [Bacillus cereus HuB5-5] (ZP_17525503.1)	55 in 94 aa
279	127	-	hypothetical protein	hypothetical protein ICU_04017 [Bacillus cereus BAG2X1-1] (ZP_17375524.1)	41 in 122
280	226	-	IS231-like transposase	Transposase for insertion sequence element IS231B [Bacillus mycoides Rock1-4] (ZP_04166736.1)	91 in 226
281	347	+	pseudogene		
282	93	-	pseudogene		
283	93	-	pseudogene		
284	119	-	copper amine oxidase-like domain-containing protein	hypothetical protein IIM_05100 [Bacillus cereus VD107] (ZP_17590246.1)	77 in 119 aa

285	84	-	transporter	Transporter [Bacillus mycoides Rock3-17] (ZP_04155868.1)	84 in 83 aa
286	45	-	hypothetical protein	hypothetical protein IEI_02611 [Bacillus cereus BAG5X2-1] (ZP_17436268.1)	70 in 56
287	340	-	mosquitocidal toxin gene	hypothetical protein bthur0009_54170 [Bacillus thuringiensis serovar andalousiensis BGSC 4AW1] (ZP_04099749.1)	32 in 341
288	478	+	IS231-like transposase	transposase for insertion sequence element IS231F, partial [Bacillus cereus VDM022] (ZP_17640768.1)	91 in 478 aa
289	226	-	IS231-like transposase	Transposase for insertion sequence element IS231B [Bacillus mycoides Rock1-4] (ZP_04166736.1)	91 in 226
290	293	-	transposase	Transposase, IS204/IS1001/IS1096/IS1165 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149049.1)	84 in 120 aa
291	127	-	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89 aa
292	144	-	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1|)	45 in 141 aa
293	708	-	Penicillin-binding protein transpeptidase	hypothetical protein IEE_02638 [Bacillus cereus BAG5X1-1] (ZP_17430747.1)	85 in 707 aa
294	235	-	peptidase M15B and M15C DD-carboxypeptidase VanY/endolysin	hypothetical protein ICG_05453 [Bacillus cereus BAG1X1-3] (ZP_17370831.1)	80 in 235
295	298	-	Serine-type D-Ala-D-Ala carboxypeptidase	hypothetical protein ICG_05452 [Bacillus cereus BAG1X1-3] (ZP_17370830.1)	78 in 298
296	384	-	sensor histidine kinase VanS	hypothetical protein ICG_05451 [Bacillus cereus BAG1X1-3] (ZP_17370829.1)	77 in 384
297	235	-	two-component response regulator VanR	hypothetical protein IC3_04873 [Bacillus cereus VD142] (ZP_17347204.1)	83 in 235
298	171	-	Invasion protein IagB domain protein	hypothetical protein IEQ_02103 [Bacillus cereus BAG6X1-2] (ZP_17459015.1)	81 in 171
299	278	-	Peptidoglycan N-acetylglucosamine deacetylase	hypothetical protein ICG_05448 [Bacillus cereus BAG1X1-3] (ZP_17370826.1)	82 in 278
300	57	-	hypothetical protein	hypothetical protein IIM_01620 [Bacillus cereus VD107] (ZP_17586766.1)	75 in 51
301	208	-	IS3 family transposase orfB	Integrase [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149206.1)	87 in 208

302	300	-	ISL3 family transposase	hypothetical protein [Bacillus thuringiensis] (WP_000098666.1)	87 in 167 aa
303	182	-	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89
304	144	-	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	45 in 141
305	431	+	IS232 transposase-like protein IstA	transposase for insertion sequence element IS232 [Bacillus cereus VD154] (ZP_17603282.1)	91 in 431 aa
306	86	+	IS232 transposase-like protein IstB	Insertion sequence IS232 putative ATP-binding protein [Bacillus thuringiensis serovar kurstaki str. T03a001] (ZP_04117980.1)	97 in 86
307	378	-	Transposase IS116/IS110/IS902	hypothetical protein IIA_05308 [Bacillus cereus VD014] (ZP_17559904.1)	76 in 403
308	110	-	transposase of ISAar40, IS3 family, IS3 group, orfA	transposase of ISAar40, IS3 family, IS3 group, orfA [Bacillus thuringiensis serovar finitimus YBT-020]	88 in 78 aa
309	260	-	Mono-ADP-ribosyltransferase C3	hypothetical protein IC3_04863 [Bacillus cereus VD142] (ZP_17347194.1)	43 in 251
310	321	-	Serine-type D-Ala-D-Ala carboxypeptidase	Serine-type D-Ala-D-Ala carboxypeptidase [Bacillus pseudogenemycoides DSM 12442] (ZP_04150951.1)	59 in 329
311	333	-	amino-acid racemase	hypothetical protein ICG_05446 [Bacillus cereus BAG1X1-3] (ZP_17370824.1)	86 in 333
312	360	-	alanine racemase domain-containing protein	hypothetical protein ICG_05445 [Bacillus cereus BAG1X1-3] (ZP_17370823.1)	84 in 360 aa
313	457	-	UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase	UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase [Bacillus cereus VD107] (ZP_17586763.1)	85 in 457
314	357	-	D-alanine-D-alanine ligase	D-alanine-D-alanine ligase [Bacillus cereus BAG1X1-3] (ZP_17370821.1)	83 in 357
315	256	-	pseudogene		
316	234	-	pseudogene		
317	112	-	transposase of ISAar40, IS3 family, IS3 group, orfA	transposase of ISAar40, IS3 family, IS3 group, orfA [Bacillus thuringiensis serovar finitimus YBT-020] (YP_005563640.1)	90 in 81 aa

318	144	+	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	55 in 141 aa
319	127	+	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89 aa
320	293	+	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149049.1)	84 in 120 aa
321	255	+	pseudogene		
322	299	-	35.8-kilodalton mosquitocidal toxin	hypothetical protein bthur0007_54850 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111632.1)	51 in 300 aa
323	249	-	hypothetical protein	hypothetical protein bthur0014_51840 [Bacillus thuringiensis IBL 4222] (ZP_04068138.1)	96 in 103 aa
324	163	-	pseudogene		
325	426	-	Collagen triple helix repeat-containing protein	hypothetical protein BCK_12595 [Bacillus cereus FRI-35] (YP_006596620.1)	45 in 410
326	294	-	pseudogene		
327	101	-	transposase	hypothetical protein ICE_05215 [Bacillus cereus BAG1X1-2] (ZP_17364725.1)	99 in 101 aa
328	256	+	mosquitocidal toxin gene	hypothetical protein bthur0012_54310 [Bacillus thuringiensis serovar pulsiensis BGSC 4CC1] (ZP_04081751.1)	30 in 259 aa
329	207	-	pseudogene		
330	403	-	Transposase IS116/IS110/IS902	hypothetical protein IIA_05308 [Bacillus cereus VD014] (ZP_17559904.1)	81 in 403 aa
331	66	-	pseudogene		
332	95	-	Phage protein	hypothetical protein bthur0014_3830 [Bacillus thuringiensis IBL 4222] (ZP_04063427.1)	87 in 95 aa
333	81	-	Phage protein	hypothetical protein II5_04377 [Bacillus cereus MSX-A1] (ZP_17541249.1)	94 in 81 aa
334	175	-	pseudogene		

335	79	-	pseudogene		
336	324	+	pseudogene		
337	93	-	hypothetical protein	hypothetical protein II5_04381 [Bacillus cereus MSX-A1] (ZP_17541253.10)	88 in 93 aa
338	95	-	pseudogene		
339	385	-	pseudogene		
340	190	+	cryBP1 family protein	hypothetical protein MC28_E061 [Bacillus thuringiensis MC28] (YP_006815559.1)	63 in 152 aa
341	411	-	pseudogene		
342	75	-	hypothetical protein	hypothetical protein MC28_E058 [Bacillus thuringiensis MC28] (YP_006815556.1)	79 in 75 aa
343	97	-	hypothetical protein	hypothetical protein MC28_E062 [Bacillus thuringiensis MC28] (YP_006815560.1)	96 in 68 aa
344	424	-	Retron-type reverse transcriptase	hypothetical protein bthur0005_3190 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04118566.1)	95 in 424 aa
345	431	+	IS232 family transposase IstA	transposase for insertion sequence element IS232 [Bacillus cereus VD154] (ZP_17603282.1)	91 in 431 aa
346	250	+	IS232 family transposase IstB	hypothetical protein IK5_00386 [Bacillus cereus VD154] (ZP_17603283.1)	98 in 250 aa
347	157	+	pseudogene		
348	722	+	Cry20Bb1	Cry20-like delta endotoxin [Bacillus thuringiensis] (ACS93601.1)	73 in 766 aa
349	187	-	pseudogene		
350	272	-	pseudogene		
351	201	-	pseudogene		
352	375	-	spore germination protein GerZC	putative spore germination protein [Bacillus thuringiensis serovar israelensis] (YP_001573812.1)	85 in 375 aa
353	367	-	spore germination protein GerZB	putative spore germination protein [Bacillus thuringiensis serovar israelensis] (YP_001573813.1)	90 in 367 aa

354	491	-	spore germination protein GerZA	putative spore germination protein [Bacillus thuringiensis serovar israelensis] (YP_001573814.1)	95 in 490 aa
355	190	-	pseudogene		
356	194	-	Cry20-like delta endotoxin	hypothetical protein MC28_D170 [Bacillus thuringiensis MC28] (YP_006815458.1)	50 in 212 aa
357	456	-	Calcineurin-like phosphoesterase	hemagglutinin-related protein [Bacillus thuringiensis serovar israelensis] (YP_001573868.1)	52 in 456
358	85	-	hypothetical protein	hypothetical protein MC28_D172 [Bacillus thuringiensis MC28] (YP_006815460.1)	82 in 85 aa
359	128	-	pseudogene		
360	253	-	pseudogene	(DNA topoisomerase III, 146kbp fragment insertion site)	
361	121	-	DNA topoisomerase TopB	DNA topoisomerase 3 [Bacillus cereus G9842] (YP_002454834.1)	98 in 121 aa
362	60	-	pseudogene		
363	304	-	ATPase AAA	stage V sporulation protein K [Bacillus cereus G9842] (YP_002454835.1)	99 in 304 aa
364	241	-	ATPase AAA+	stage V sporulation protein K [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111626.1)	95 in 239
365	140	-	hypothetical protein	hypothetical protein bthur0004_60870 [Bacillus thuringiensis serovar sotto str. T04001] (ZP_04130236.1)	96 in 140 aa
366		-	hypothetical protein	hypothetical protein BCG9842_0172 [Bacillus cereus G9842] (YP_002454837.1)	97 in 338 aa
367	458	+	nlpC/P60 family protein	hypothetical protein IK9_05470 [Bacillus cereus VD166] (ZP_17621143.1)	98 in 458 aa
368	216	-	hypothetical protein	hypothetical protein bthur0001_53860 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148816.1)	95 in 216 aa
369	1048	-	cell wall endopeptidase, family M23/M37	cell wall endopeptidase, family M23/M37 [Bacillus cereus G9842] (YP_002454840.1)	99 in 1048 aa
370	1562	-	Reticulocyte binding protein	hypothetical protein BCG9842_0176 [Bacillus cereus G9842] (YP_002454841.1)	99 in 1562 aa

371	698	-	hypothetical protein	hypothetical protein bthur0007_54690 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111616.1)	99 in 698 aa
372	301	-	hypothetical protein	hypothetical protein IK9_05465 [Bacillus cereus VD166] (ZP_17621138.1)	99 in 301 aa
373	108	-	hypothetical protein	hypothetical protein BCG9842_0179 [Bacillus cereus G9842] (YP_002454844.1)	97 in 108 aa
374	216	-	hypothetical protein	hypothetical protein bthur0007_54660 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111613.1)	99 in 216 aa
375	142	-	hypothetical protein	hypothetical protein BCG9842_0181 [Bacillus cereus G9842] (YP_002454846.1)	99 in 142 aa
376	96	+	ATP synthase F0 subunit 6	hypothetical protein BCG9842_0182 [Bacillus cereus G9842] (YP_002454847.1)	98 in 96 aa
377	433	-	DNA primase	putative DNA primase [Bacillus cereus G9842] (YP_002454848.1)	98 in 433 aa
378	551	-	hypothetical protein	hypothetical protein BCG9842_0184 [Bacillus cereus G9842] (YP_002454849.1)	97 in 551 aa
379	85	-	hypothetical protein	hypothetical protein BCG9842_0185 [Bacillus cereus G9842] (YP_002454850.1)	92 in 73 aa
380	379	+	DNA polymerase III subunit beta	DNA polymerase III, beta subunit [Bacillus cereus G9842] (YP_002454851.1)	99 in 379 aa
381	137	+	hypothetical protein	hypothetical protein BCG9842_0187 [Bacillus cereus G9842] (YP_002454852.1)	99 in 137 aa
382	178	+	hypothetical protein	hypothetical protein BCG9842_0188 [Bacillus cereus G9842] (YP_002454853.1)	99 in 178 aa
383	227	+	hypothetical protein	hypothetical protein BCG9842_0189 [Bacillus cereus G9842] (YP_002454854.1)	97 in 227 aa
384	406	+	hypothetical protein	hypothetical protein BCG9842_0190 [Bacillus cereus G9842] (YP_002454855.1)	96 in 406 aa
385	307	+	ThiF family protein	ThiF family protein [Bacillus cereus G9842] (YP_002454856.1)	99 in 307 aa
386	337	-	hypothetical protein	hypothetical protein BCG9842_0192 [Bacillus cereus G9842] (YP_002454857.1)	98 in 337 aa
387	234	-	hypothetical protein	hypothetical protein BCG9842_0193 [Bacillus cereus G9842] (YP_002454858.1)	98 in 222 aa
388	136	-	hypothetical protein	hypothetical protein BCG9842_0194 [Bacillus cereus G9842] (YP_002454859.1)	100 in 136 aa
389	83	-	hypothetical protein	hypothetical protein BCG9842_0195 [Bacillus cereus G9842] (YP_002454860.1)	100 in 83 aa
390	81	-	hypothetical protein	hypothetical protein bthur0007_54480 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111595.1)	100 in 81 aa
391	81	-	hypothetical protein	hypothetical protein BCG9842_0197 [Bacillus cereus G9842] (YP_002454862.1)	100 in 81 aa

392	308	-	hypothetical protein	hypothetical protein bthur0007_54460 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111593.1)	97 in 308 aa
393	327	-	hypothetical protein	conserved membrane protein, putative [Bacillus cereus G9842] (YP_002454864.1)	98 in 327 aa
394	498	-	bacterial type II/IV secretion system protein	bacterial type II/IV secretion system protein [Bacillus cereus G9842] (YP_002454865.1)	97 in 498 aa
395	272	-	SAF domain family protein	hypothetical protein BCG9842_0201 [Bacillus cereus G9842] (YP_002454866.1)	98 in 272 aa
396	303	-	flp pilus assembly protein CpaB	hypothetical secreted protein [Bacillus cereus G9842] (YP_002454867.1)	99 in 303 aa
397	159	-	hypothetical protein	hypothetical protein BCG9842_0203 [Bacillus cereus G9842] (YP_002454868.1)	100 in 159 aa
398	863	-	PQQ enzyme repeat domain protein	hypothetical protein BCG9842_0204 [Bacillus cereus G9842] (YP_002454869.1)	99 in 863 aa
399	242	-	hypothetical protein	hypothetical protein BCG9842_0205 [Bacillus cereus G9842] (YP_002454870.1)	96 in 248 aa
400	45	-	hypothetical protein	hypothetical protein BCG9842_0206 [Bacillus cereus G9842] (YP_002454871.1)	91 in 45 aa
401	214	-	hypothetical protein	hypothetical protein BCG9842_0207 [Bacillus cereus G9842] (YP_002454872.1)	95 in 213 aa
402	182	-	hypothetical protein	conserved hypothetical membrane protein, putative [Bacillus cereus G9842] (YP_002454873.1)	100 in 182 aa
403	507	-	FtsZ/tubulin-related protein	putative FtsZ/tubulin-related protein [Bacillus cereus G9842] (YP_002454874.1)	99 in 507 aa
404	142	-	hypothetical protein	hypothetical protein BCG9842_0210 [Bacillus cereus G9842] (YP_002454875.1)	99 in 142 aa
405	51	-	hypothetical protein	ypothetical protein BCG9842_0211 [Bacillus cereus G9842] (YP_002454876.1)	94 in 48 aa
406	509	+	hypothetical protein	hypothetical protein BCG9842_0213 [Bacillus cereus G9842] (YP_002454878.1)	99 in 509 aa
407	962	+	TraG/TraD family conjugation protein	hypothetical protein IK9_05431 [Bacillus cereus VD166] (ZP_17621104.1)	94 in 962 aa
408	447	-	replicative DNA helicase (ori-related genes)	replicative DNA helicase [Bacillus cereus G9842] (YP_002454881.1)	99 in 447 aa
409	178	+	RsfA family transcription factor (ori-related genes)	transcription factor, RsfA family [Bacillus cereus G9842] (YP_002454882.1)	96 in 178 aa

410	242	-	MerR family transcriptional regulator (ori-related genes)	hypothetical protein BCG9842_0218 [Bacillus cereus G9842] (YP_002454883.1)	99 in 242 aa

Table S2. The CDS in pMOGI222 and their annotations.

CDS	size (aa)	Strand	annotation	Best hit in databases (GenBank no.)	(% aa identity)
1	129	-	hypothetical protein (ori-related genes)	hypothetical protein IIM_05167 [Bacillus cereus VD107] (ZP_17590313.1)	80 in 129 aa
2	275	-	chromosome partitioning ATPase (ori-related genes)	ATPase [Bacillus cereus] (WP_000335378.1)	90 in 269 aa
3	518	+	replication initiation protein (ori-related genes)	replication initiation protein [Bacillus cereus] (WP_001099049.1)	97 in 518 aa
4	88	+	pseudogene		
5	710	-	Tn3 family transposase	hypothetical protein IIO_06154 [Bacillus cereus VD115] (ZP_17596662.1)	75 in 667 aa
6	284	+	TnpI resolvase	TnpI resolvase [Bacillus thuringiensis] (YP_001485222.1)	92 in 284 aa
7	987	+	transposase for Tn1546	transposase [Bacillus cereus] (YP_001966662.1)	98 in 987 aa
8	478	+	transposase for insertion sequence element IS231B	transposase for insertion sequence element IS231B [Bacillus thuringiensis MC28] (YP_006815473.1)	95 in 478 aa
9	169	+	pseudogene		
10	215	-	DNA integration/recombination/ inversion protein	DNA integration/recombination/inversion protein [Bacillus thuringiensis IBL 4222] (ZP_04069695.1)	96 in 209 aa
11	74	+	hypothetical protein	hypothetical protein IKG_05542 [Bacillus cereus VD200] (ZP_17633897.1)	81 in 75 aa
12	118	+	hypothetical protein	hypothetical protein pBt066 [Bacillus thuringiensis serovar israelensis] (YP_001573800.1)	94 in 118 aa
13	49	+	hypothetical protein	hypothetical protein IKM_05534 [Bacillus cereus VDM022] (ZP_17640732.1)	78 in 49 aa
14	305	-	chromosome segregation ATPase	hypothetical protein bthur0005_53640 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123497.1)	90 in 309 aa

15	102	-	hypothetical protein	hypothetical protein MC28_C058 [Bacillus thuringiensis MC28] (YP_006815245.1)	95 in 102 aa
16	96	-	ArsR family transcriptional regulator	hypothetical protein IC1_05970 [Bacillus cereus VD022] (ZP_17341493.1)	82 in 94 aa
17	92	+	small DNA-binding protein	small DNA-binding protein [Bacillus thuringiensis serovar israelensis] (YP_001573818.1)	91 in 92 aa
18	57	-	RNA chaperone Hfq	RNA chaperone Hfq [Bacillus thuringiensis IBL 4222] (ZP_04069280.1)	98 in 52 aa
19	93	-	AbrB family transcriptional regulator	putative transcriptional regulator [Bacillus cereus Q1] (YP_002533363.1)	84 in 93 aa
20	61	+	hypothetical protein	hypothetical protein pBt095 [Bacillus thuringiensis serovar israelensis] (YP_001573821.1)	80 in 61 aa
21	260	-	CAAX amino terminal protease family protein	hypothetical protein II3_05177 [Bacillus cereus MC67] (ZP_17536275.1)	37 in 240 aa
22	710	-	DNA topoisomerase III	DNA topoisomerase III [Bacillus cereus VD154] (ZP_17609214.1)	96 in 445 aa
23	219	-	hypothetical protein	hypothetical protein bthur0005_53750 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123508.1)	98 in 219 aa
24	102	-	hypothetical protein	hypothetical protein bthur0005_53760 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123509.1)	100 in 102 aa
25	269	-	hypothetical protein	hypothetical protein bthur0005_53770 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123510.1)	99 in 269 aa
26	53	-	hypothetical protein	hypothetical protein IK5_06321, partial [Bacillus cereus VD154] (ZP_17609218.1)	100 in 53 aa
27	81	-	hypothetical protein	hypothetical protein MC28_D003 [Bacillus thuringiensis MC28] (YP_006815291.1)	89 in 81 aa
28	178	-	hypothetical protein	hypothetical protein IK5_05725 [Bacillus cereus VD154] (ZP_17608622.1)	97 in 178 aa
29	416	-	hypothetical protein	hypothetical protein IK5_05727 [Bacillus cereus VD154] (ZP_17608624.1)	96 in 416 aa

30	195	-	hypothetical protein	hypothetical protein IK5_05728 [Bacillus cereus VD154] (ZP_17608625.1)	94 in 195 aa
31	342	-	conjugation protein TraL	hypothetical protein IK5_05729 [Bacillus cereus VD154] (ZP_17608626.1)	99 in 342 aa
32	304	-	hypothetical protein	hypothetical protein bthur0005_60130 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04124026.1)	100 in 304 aa
33	301	-	hypothetical protein	hypothetical protein bthur0005_60120 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04124025.1)	98 in 301 aa
34	765	-	conjugation protein TrsK	hypothetical protein IK5_05732, partial [Bacillus cereus VD154] (ZP_17608629.1)	95 in 393 aa
35	605	-	conjugal transfer protein TraE	protein TrsE [Bacillus sp. GeD10] (CCW06843.1)	99 in 605 aa
36	224	-	hypothetical protein	hypothetical protein IK5_05737 [Bacillus cereus VD154] (ZP_17608634.1)	97 in 224 aa
37	106	-	hypothetical protein	hypothetical protein IK5_05738 [Bacillus cereus VD154] (ZP_17608635.1)	100 in 106 aa
38	117	-	hypothetical protein	hypothetical protein IK5_05739 [Bacillus cereus VD154] (ZP_17608636.1)	100 in 117 aa
39	58	-	hypothetical protein	hypothetical protein IK5_05740 [Bacillus cereus VD154] (ZP_17608637.1)	97 in 58 aa
40	504	-	Surface adhesion protein; Bacillolysin / Insecticidal delta-endotoxin protein	hypothetical protein IK5_05741 [Bacillus cereus VD154] (ZP_17608638.1)	96 in 505 aa
41	242	-	hypothetical protein	hypothetical protein bthur0005_57990 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123853.1)	97 in 242 aa
42	81	-	hypothetical protein	hypothetical protein EBGED10_35650 [Bacillus sp. GeD10] (CCW06835.1)	93 in 81 aa
43	154	-	TRAG family protein	hypothetical protein bthur0005_57970 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123851.1)	88 in 154 aa
44	153	-	hypothetical protein	hypothetical protein bthur0005_58140 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123864.1)	92 in 153 aa
45	751	-	TQXA domain-containing protein	TQXA domain-containing protein [Bacillus cereus VDM062] (ZP_17653212.1)	64 in 747 aa

46	69	-	Xre family transcriptional regulator	Transcriptional regulator, Xre [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123550.1)	96 in 69 aa
47	119	+	MerR family transcriptional regulator	hypothetical protein IK5_06346 [Bacillus cereus VD154] (ZP_17609243.1)	100 in 119 aa
48	409	+	hypothetical protein	hypothetical protein IK5_05988 [Bacillus cereus VD154] (ZP_17608885.1)	87 in 410 aa
49	166	+	hypothetical protein	hypothetical protein III_06028 [Bacillus cereus VD078] (ZP_17579226.1)	77 in 176 aa
50	408	+	hypothetical protein	hypothetical protein III_05946 [Bacillus cereus VD078] (ZP_17579144.1)	98 in 408 aa
51	149	+	hypothetical protein	hypothetical protein IK5_06231 [Bacillus cereus VD154] (ZP_17609128.1)	91 in 149 aa
52	71	+	hypothetical protein	hypothetical protein bthur0005_51970 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123363.1)	90 in 60 aa
53	80	-	hypothetical protein	hypothetical protein bthur0005_52320 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123394.1)	93 in 43 aa
54	98	-	hypothetical protein	hypothetical protein IK5_05757 [Bacillus cereus VD154] (ZP_17608654.1)	96 in 98 aa
55	109	-	hypothetical protein	hypothetical protein IK5_05758 [Bacillus cereus VD154] (ZP_17608655.1)	98 in 109 aa
56	86	-	hypothetical protein	hypothetical protein IK5_05759 [Bacillus cereus VD154] (ZP_17608656.1)	98 in 86 aa
57	949	+	MobA/MobL family protein	hypothetical protein IK5_05760 [Bacillus cereus VD154] (ZP_17608657.1)	98 in 949 aa
58	216	+	hypothetical protein	hypothetical protein bthur0005_60700 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04124077.1)	96 in 216 aa
59	48	-	hypothetical protein	hypothetical protein IK5_05762 [Bacillus cereus VD154] (ZP_17608659.1)	100 in 48 aa
60	348	-	response regulator aspartate phosphatase I	Prophage LambdaBa01, TPR domain protein [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04124075.1)	98 in 348 aa
61	98	+	hypothetical protein	hypothetical protein bthur0005_60670 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04124074.1)	98 in 98 aa
62	61	+	hypothetical protein	hypothetical protein IK5_06005 [Bacillus cereus VD154] (ZP_17608902.1)	97 in 61 aa
63	393	+	FtsK/SpoIIIE ATPase	hypothetical protein IK5_05766 [Bacillus cereus VD154] (ZP_17608663.1)	94 in 393 aa

64	205	+	Phage protein	hypothetical protein IK5_05767 [Bacillus cereus VD154] (ZP_17608664.1)	94 in 204 aa
65	75	-	XRE family transcriptional regulator	conserved hypothetical protein [Geobacillus sp. G11MC16] (ZP_03149500.1)	65 in 74 aa
66	72	-	hypothetical protein	hypothetical protein IK5_06243 [Bacillus cereus VD154] (ZP_17609140.1)	81 in 73 aa
67	56	-	hypothetical protein	hypothetical protein EBGED10_14800 [Bacillus sp. GeD10] (CCW04762.1)	73 in 56 aa
68	79	-	hypothetical protein	hypothetical protein IK5_06245 [Bacillus cereus VD154] (ZP_17609142.1)	85 in 79 aa
69	78	-	hypothetical protein	hypothetical protein bmyco0002_58430 [Bacillus mycoides Rock1-4] (ZP_04166458.1)	42 in 79 aa
70	82	-	hypothetical protein	hypothetical protein III_05964 [Bacillus cereus VD078] (ZP_17579162.1)	75 in 81 aa
71	183	-	hypothetical protein	hypothetical protein bthur0005_55920 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123674.1)	89 in 183 aa
72	157	+	hypothetical protein	hypothetical protein [Bacillus cereus](WP_016099623.1)	99 in 157 aa
73	414	-	tetratricopeptide repeat family protein	hypothetical protein bmyco0002_59880 [Bacillus mycoides Rock1-4] (ZP_04166587.1)	89 in 404 aa
74	130	-	TIR_2 superfamily protein	hypothetical protein bmyco0002_59870 [Bacillus mycoides Rock1-4] (ZP_04166586.1)	94 in 130 aa
75	61	-	pseudogene		
76	234	+	hypothetical protein	conserved hypothetical protein [Bacillus cereus 03BB108] (ZP_03112632.1)	92 in 234 aa
77	126	-	hypothetical protein	hypothetical protein RBTH_07022 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00739166.1)	74 in 125 aa
78	68	-	Cro/CI family transcriptional regulator	Transcriptional regulator, Cro/CI family [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00739165.1)	85 in 68 aa
79	298	+	Secreted subtilisin-like serine protease	protease [Bacillus cereus 03BB108] (ZP_03112649.1)	80 in 286 aa
80	96	+	Thiol-disulfide oxidoreductase BdbC	Thiol-disulfide oxidoreductase BdbC [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00739163.1)	84 in 96 aa

81	153	+	Thioredoxin	Thioredoxin [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00739162.1)	84 in 153 aa
82	71	-	hypothetical protein	hypothetical protein IK5_06243 [Bacillus cereus VD154] (ZP_17609140.1)	63 in 73 aa
83	171	-	hypothetical protein	hypothetical protein MC28_E153 [Bacillus thuringiensis MC28] (YP_006815651.1)	85 in 132 aa
84	99	-	hypothetical protein	hypothetical protein MC28_E152 [Bacillus thuringiensis MC28] (YP_006815650.1)	95 in 99 aa
85	49	+	hypothetical protein	hypothetical protein MC28_E151 [Bacillus thuringiensis MC28] (YP_006815649.1)	96 in 49 aa
86	118	+	hypothetical protein	hypothetical protein MC28_E150 [Bacillus thuringiensis MC28] (YP_006815648.1)	90 in 120 aa
87	148	-	hypothetical protein	hypothetical protein MC28_E149 [Bacillus thuringiensis MC28] (YP_006815647.1)	73 in 158 aa
88	190	+	hypothetical protein	hypothetical protein MC28_E147 [Bacillus thuringiensis MC28](YP_006815645.1)	81 in 190 aa
89	56	+	pseudogene		
90	144	+	hypothetical protein	hypothetical protein BTF1_31851 [Bacillus thuringiensis HD-789] (YP_006613948.1)	100 in 144 aa
91	131	+	hypothetical protein	hypothetical protein RBTH_07781 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00738434.1)	100 in 131 aa
92	137	+	Cobalamin synthesis protein P47K	hypothetical protein RBTH_07780 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00738433.1)	85 in 137 aa
93	390	+	lipase	lipase [Bacillus thermoamylovorans] (BAH70300.1)	57 in 375 aa
94	284	+	TnpI resolvase	TnpI resolvase [Bacillus thuringiensis] (YP_001485222.1)	92 in 284 aa

95	987	+	transposase for transposon Tn1546	transposase [Bacillus cereus] (YP_001966662.1)	98 in 987 aa
96	53	+	hypothetical protein	hypothetical protein bthur0013_55230 [Bacillus thuringiensis IBL 200] (ZP_04075187.1)	72 in 53 aa
97	260	+	hypothetical protein	hypothetical protein BSSC8_22580 [Bacillus subtilis subsp. subtilis str. SC-8] (ZP_12671314.1)	32 in 230 aa
98	111	+	hypothetical protein	hypothetical protein bpmyx0001_50440 [Bacillus pseudomycoides DSM 12442] (ZP_04154217.1)	84 in 58 aa
99	172	+	chromosome segregation ATPase	hypothetical protein IK5_06171 [Bacillus cereus VD154] (ZP_17609068.1)	91 in 176 aa
100	194	-	DNA-Invertase BINR	hypothetical protein IKM_05568 [Bacillus cereus VDM022] (ZP_17640766.1)	83 in 187 aa
101	74	-	hypothetical protein	hypothetical protein SSIL_2306 [Solibacillus silvestris StLB046] (YP_006462875.1)	68 in 74 aa
102	199	-	hypothetical protein	hypothetical protein BN424_3862 [Carnobacterium maltaromaticum LMA28] (YP_006994581.1)	67 in 183 aa
103	313	-	epoxide hydrolase 2	epoxide hydrolase 2 [Microscilla marina ATCC 23134] (ZP_01690851.1)	83 in 307 aa
104	309	-	DeoR family transcriptional regulator	hypothetical protein IYC_05053 [Clostridium sporogenes PA 3679] (ZP_18252214.1)	47 in 310 aa
105	183	+	Resolvase	Resolvase [Bacillus thuringiensis IBL 200] (ZP_04075383.1)	97 in 182 aa
106	220	-	mosquitocidal toxin	RecName: Full=Pesticidal crystal protein cry19Ba; AltName: Full=78 kDa crystal protein; AltName: Full=Crystaline entomocidal protoxin; AltName: Full=Insecticidal delta-endotoxin CryXIXB(a) (O86170.1)	97 in 217 aa
107	139	+	pseudogene		
108	825	+	Cry27Ab1	RecName: Full=Pesticidal crystal protein cry27Aa; AltName: Full=94 kDa crystal protein; AltName: Full=Crystaline entomocidal protoxin; AltName: Full=Insecticidal delta-endotoxin CryXXVIIA(a) (Q9S597.1)	83 in 829 aa

109	160	+	group-specific protein	hypothetical protein IE5_00813 [Bacillus cereus BAG3X2-2] (ZP_17400155.1)	79 in 140 aa
110	261	+	response regulator aspartate phosphatase K	response regulator aspartate phosphatase [Bacillus thuringiensis serovar thuringiensis str. IS5056] (YP_007491988.1)	75 in 238 aa
111	478	-	transposase for insertion sequence element IS231B	transposase for insertion sequence element IS231B [Bacillus thuringiensis MC28] (YP_006815473.1)	95 in478 aa
112	260	-	hypothetical protein	hypothetical protein BSSC8_22580 [Bacillus subtilis subsp. subtilis str. SC-8] (ZP_12671314.1)	32 in 230 aa
113	53	-	hypothetical protein	hypothetical protein bthur0013_55230 [Bacillus thuringiensis IBL 200] (ZP_04075187.1)	72 in 53 aa
114	65	+	pseudogene		
115	162	-	copper amine oxidase domain-containing protein	Hypothetical protein RBTH_08755 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00741906.1)	77 in 162 aa
116	123		pseudogene		
117	460	-	Collagen-like triple helix repeat protein	Collagen-like triple helix repeat protein [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00741909.1)	60 in 359 aa
118	248	-	IS605 OrfB family transposase	transposase, OrfB family [Bacillus cereus AH1134] (ZP_03233746.1)	94 in 247 aa
119	473	-	Transposase for insertion sequence element IS231B	Transposase for insertion sequence element IS231B [Bacillus thuringiensis serovar andalousiensis BGSC 4AW1] (ZP_04099942.1)	92 in 472 aa
120	86	+	DNA binding protein	DNA binding protein [Bacillus thuringiensis serovar andalousiensis BGSC 4AW1] (ZP_04099943.1)	91 in 86 aa
121	144	+	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	45 in 141 aa
122	127	+	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89 aa

123	293	+	Transposase, IS204/IS1001/IS1096/IS1165	Transposase, IS204/IS1001/IS1096/IS1165 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149049.1)	84 in 120 aa
124	64	+	Transposase for insertion sequence element IS231B	transposase for insertion sequence element IS231E, partial [Bacillus cereus BAG6O-1] (ZP_17445153.1)	94 in 64 aa
125	115	+	Exosporium protein ExsB	hypothetical protein IIA_02388 [Bacillus cereus VD014] (ZP_17556984.1)	70 in 126 aa
126	162	-	copper amine oxidase domain-containing protein	Hypothetical protein RBTH_08755 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00741906.1)	78 in 162 aa
127	111	-	Transporter	Transporter [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00741907.1)	79 in 120 aa
128	201	-	Collagen-like triple helix repeat protein	Collagen-like triple helix repeat protein [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00741909.1)	64 in 216 aa
129	235	+	pseudogene		
130	137	-	Ice nucleation protein	putative deletion pseudogene prduct [Bacillus thuringiensis serovar israelensis] (YP_001573797.1)	61 in 135 aa
131	471	-	Calcineurin-like phosphoesterase; mosquitocidal toxin protein	hemagglutinin-related protein [Bacillus thuringiensis serovar israelensis] (YP_001573868.1)	48 in 483 aa
132	185	+	19kda accessory protein	hypothetical protein MC28_E061 [Bacillus thuringiensis MC28] (YP_006815559.1)	51 in 185 aa
133	683	+	pesticidial crystal protein cry4AA	pesticidial crystal protein cry4AA [Bacillus thuringiensis serovar finitimus YBT-020] (YP_005569289.1)	30 in 648 aa
134	506	-	crystal protein ET69	crystal protein ET69 [Bacillus thuringiensis] (Sequence ID: gb|AAK64558.1)	33 in 480 aa
135	77	-	hypothetical protein	hypothetical protein MC28_E058 [Bacillus thuringiensis MC28] (YP_006815556.1)	85 in 73 aa

136	98	-	hypothetical protein	hypothetical protein MC28_E062 [Bacillus thuringiensis MC28] (YP_006815560.1)	97 in 67 aa
137	91	-	pseudogene		
138	144	+	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	45 in 141 aa
139	127	+	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89 aa
140	293	+	pseudogene		
141	288	-	pseudogene		
142	86	-	DNA binding protein	DNA binding protein [Bacillus thuringiensis serovar andalousiensis BGSC 4AW1] (ZP_04099943.1)	91 in 86 aa
143	473	+	Transposase for insertion sequence element IS231B	Transposase for insertion sequence element IS231B [Bacillus thuringiensis serovar andalousiensis BGSC 4AW1] (ZP_04099942.1)	92 in 472 aa
144	286	+	hypothetical protein	hypothetical protein RBTH_07849 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00738488.1)	98 in 286 aa
145	352	+	hypothetical protein	hypothetical protein pBt074 [Bacillus thuringiensis serovar israelensis] (YP_001573807.1)	90 in 352 aa
146	248	+	hypothetical protein	hypothetical protein bthur0014_62240 [Bacillus thuringiensis IBL 4222] (ZP_04069141.1)	88 in 248 aa
147	92	-	DNA-binding protein HU	small DNA-binding protein [Bacillus thuringiensis serovar israelensis] (YP_001573818.1)	97 in 92 aa
148	48	+	hypothetical protein	hypothetical protein IC1_05970 [Bacillus cereus VD022] (ZP_17341493.1)	67 in 48 aa
149	89	-	transposase of ISAar40, IS3 family, IS3 group, orfA	transposase of ISAar40, IS3 family, IS3 group, orfA [Bacillus thuringiensis serovar finitimus YBT-020] (YP_005563640.1)	85 in 88 aa

150	70	-	Methionine-rich protein	hypothetical protein bpmyx0001_29000 [Bacillus pseudomycoides DSM 12442] (ZP_04152091.1)	84 in 70 aa
151	73	+	pseudogene		
152	480	-	GntR family transcriptional regulator	GntR domain-containing protein [Bacillus thuringiensis serovar israelensis] (YP_001573828.1)	98 in 480 aa
153	299	+	GHMP kinase	kinase [Bacillus thuringiensis serovar israelensis] (YP_001573827.1)	96 in 299 aa
154	235	+	alanyl-tRNA synthetase	tRNA synthetase-related protein [Bacillus thuringiensis serovar israelensis] (YP_001573826.1)	97 in 235 aa
155	345	+	pyridoxal-phosphate dependent protein	pyridoxal-phosphate dependent protein [Bacillus thuringiensis MC28] (YP_006815619.1)	96 in 345 aa
156	394	+	class-II aminotransferase	class-II aminotransferase [Bacillus thuringiensis MC28] (YP_006815618.1)	95 in 394 aa
157	299	+	EamA-like transporter family	EamA-like transporter family [Bacillus thuringiensis MC28] (YP_006815617.1)	98 in 299 aa
158	93	+	AbrB family transcriptional regulator	putative transcriptional regulator [Bacillus thuringiensis serovar israelensis] (YP_001573820.1)	91 in 89 aa
159	65	+	RNA chaperone Hfq	RNA chaperone Hfq [Bacillus thuringiensis MC28] (YP_006815242.1)	53 in 77 aa
160	83	+	transposase of ISAar40, IS3 family, IS3 group, orfA	hypothetical protein IKM_05489 [Bacillus cereus VDM022] (ZP_17640687.1)	57 in 86 aa
161	278	-	ISPsy9, transposase OrfB	hypothetical protein IKA_05206 [Bacillus cereus VD169] (ZP_17626989.1)	99 in 278 aa
162	107	-	transposase	hypothetical protein IKA_05207 [Bacillus cereus VD169] (ZP_17626990.1)	100 in 107 aa
163	101	+	transposase	hypothetical protein ICE_05215 [Bacillus cereus BAG1X1-2] (ZP_17364725.1)	99 in 101 aa
164	144	+	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	45 in 141 aa
165	127	+	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in89 aa

166	293	+	Transposase, IS204/IS1001/IS1096/IS1165	Transposase, IS204/IS1001/IS1096/IS1165 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149049.1)	84 in120 aa
167	244	+	pseudogene		
168	195	+	pseudogene		
169	190	-	antioxidant, AhpC/TSA family	antioxidant, AhpC/TSA family [Bacillus sp. M 2-6] (ZP_10162419.1)	31 in 144 aa
170	180	+	pseudogene		
171	147	-	hypothetical protein	hypothetical protein MC28_E149 [Bacillus thuringiensis MC28] (YP_006815647.1)	81 in 123 aa
172	62	+	pseudogene		
173	82	+	hypothetical protein	hypothetical protein MC28_F175 [Bacillus thuringiensis MC28] (YP_006815858.1)	96 in 82 aa
174	386	-	DNA integration/recombination/inversion protein	DNA integration/recombination/inversion protein [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04108478.1)	95 in 386 aa
175	321	+	excisionase family DNA binding domain-containing protein	hypothetical protein IG3_06349 [Bacillus cereus HuA2-1] (ZP_17491383.1)	88 in 320 aa
176	137	-	pseudogene		
177	435	-	putative reverse transcriptase	RNA-directed DNA polymerase (Reverse transcriptase) [Bacillus thuringiensis serovar berliner ATCC 10792] (ZP_04102003.1)	99 in 435 aa
178	59	-	hypothetical protein	hypothetical protein pBMB0558_00185 [Bacillus thuringiensis CT43] (YP_004169148.1)	98 in 59 aa
179	334	-	pseudogene		
180	372	+	pseudogene		
181	59	+	hypothetical protein	hypothetical protein pBMB0558_00185 [Bacillus thuringiensis CT43] (YP_004169148.1)	98 in 59 aa

182	373	+	putative reverse transcriptase	hypothetical protein IEE_05364 [Bacillus cereus BAG5X1-1] (ZP_17433473.1)	97 in 373 aa
183	136	+	pseudogene		
184	297	+	chromosome segregation ATPase	hypothetical protein bthur0014_63540 [Bacillus thuringiensis IBL 4222] (ZP_04069259.1)	97 in 301 aa
185	116	-	hypothetical protein	hypothetical protein IKM_05468 [Bacillus cereus VDM022] (ZP_17640666.1)	99 in 116 aa
186	557	-	mosquitocidal toxin protein	hypothetical protein bthur0009_56310 [Bacillus thuringiensis serovar andalousiensis BGSC 4AW1] (ZP_04099944.1)	38 in 149 aa
187	101	+	RNA polymerase sigma factor SigX	RNA polymerase sigma factor SigX [Bacillus thuringiensis serovar israelensis] (YP_001573832.1)	93 in 72 aa
188	194	-	Transposon resolvase	hypothetical protein IKM_05559 [Bacillus cereus VDM022] (ZP_17640757.1)	99 in 194 aa
189	335	+	transposase Tn3 family protein	hypothetical protein IKM_05637, partial [Bacillus cereus VDM022] (ZP_17640835.1)	99 in 334 aa
190	431	+	transposase IstA	transposase IstA [Bacillus thuringiensis] (YP_001485232.1)	99 in 431 aa
191	250	+	ATP-binding protein IstB	ATP-binding protein IstB [Bacillus thuringiensis] (YP_001485231.1)	99 in 250 aa
192	144	+	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	45 in 141 aa
193	127	+	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89 aa
194	293	+	Transposase, IS204/IS1001/IS1096/IS1165	Transposase, IS204/IS1001/IS1096/IS1165 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149049.1)	84 in 120 aa
195	703	+	transposase Tn3 family protein	hypothetical protein IKM_05637, partial [Bacillus cereus VDM022] (ZP_17640835.1)	98 in 647 aa
196	175	+	pseudogene		
197	202	-	MerR superfamily protein	hypothetical protein IK3_05615 [Bacillus cereus VD148] (ZP_17602795.1)	78 in 466 aa

198	160	+	hypothetical protein	hypothetical protein pBT9727_0060 [Bacillus thuringiensis serovar konkukian str. 97-27] (YP_173303.1)	91 in 190 aa
199	461	-	Phosphatidylinositol-specific phospholipase	Phosphatidylinositol-specific phospholipase [Bacillus cereus AH621] (ZP_04298276.1)	78 in 466 aa
200	190	-	hypothetical protein	hypothetical protein MC28_E159 [Bacillus thuringiensis MC28] (YP_006815657.1)	91 in 190 aa
201	408	+	hypothetical protein	hypothetical protein [Bacillus cereus] (WP_016099379.1)	87 in 408 aa
202	168	-	pseudogene		
203	75	-	pseudogene		
204	293	-	Transposase, IS204/IS1001/IS1096/IS1165	Transposase, IS204/IS1001/IS1096/IS1165 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149049.1)	84 in 120 aa
205	127	-	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89 aa
206	144	-	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	45 in 141 aa
207	351	+	hypothetical protein	hypothetical protein IIE_06301 [Bacillus cereus VD045] (ZP_17566976.1)	87 in 352 aa
208	128	+	hypothetical protein	hypothetical protein IIE_06302 [Bacillus cereus VD045] (ZP_17566977.1)	54 in 127 aa
209	100	+	hypothetical protein	hypothetical protein [Bacillus cereus] (WP_016099734.1)	87 in 53 aa
210	98	+	hypothetical protein	hypothetical protein MC28_E062 [Bacillus thuringiensis MC28] (YP_006815560.1)	97 in 67 aa
211	75	+	hypothetical protein	hypothetical protein MC28_E058 [Bacillus thuringiensis MC28] (YP_006815556.1)	85 in 73 aa
212	996	-	transposase Tn3 family protein	transposase Tn3 family protein [Bacillus thuringiensis serovar finitimus YBT-020] (YP_005563651.1)	92 in 983 aa
213	280	-	integrase-recombinase protein	integrase-recombinase protein [Bacillus cereus E33L] (YP_245544.1)	90 in 280 aa

214	144	+	hypothetical protein	hypothetical protein BTG_31098 [Bacillus thuringiensis HD-771] (YP_006593824.1)	52 in 139 aa
215	54	+	hypothetical protein	hypothetical protein bthur0005_58680 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123905.1)	83 in 40 aa
216	70	+	hypothetical protein	hypothetical protein IEI_05877 [Bacillus cereus BAG5X2-1] (ZP_17439534.1)	52 in 63 aa
217	101	-	ArsR family transcriptional regulator	ArsR family transcriptional regulator [Bacillus thuringiensis serovar finitimus YBT-020] (YP_005569119.1)	68 in 94 aa
218	103	+	pseudogene		
219	68	+	hypothetical protein	hypothetical protein MC28_F146 [Bacillus thuringiensis MC28] (YP_006815829.1)	93 in 68 aa
220	239	-	hypothetical protein	hypothetical protein IEM_05053 [Bacillus cereus BAG6O-2] (ZP_17450491.1)	53 in 244 aa
221	366	+	response regulator aspartate phosphatase	hypothetical protein IC1_02796 [Bacillus cereus VD022] (ZP_17338319.1)	87 in 365 aa
222	80	+	hypothetical protein	hypothetical protein IEI_05823 [Bacillus cereus BAG5X2-1] (ZP_17439480.1)	78 in 74 aa
223	119	-	hypothetical protein	hypothetical protein IIE_06264 [Bacillus cereus VD045] (ZP_17566939.1)	93 in 118 aa
224	205	+	recombinase	hypothetical protein II5_06087 [Bacillus cereus MSX-A1] (ZP_17542959.1)	85 in 200 aa
225	56	+	pseudogene		
226	141	+	hypothetical protein	lipoprotein [Bacillus thuringiensis MC28] (YP_006815604.1)	96 in 141 aa
227	349	-	Fic family protein	Fic family protein [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00738430.1)	87 in 348 aa
228	326	-	hypothetical protein	hypothetical protein MC28_E157 [Bacillus thuringiensis MC28] (YP_006815655.1)	66 in 326 aa
229	404	-	lipase	lipase [Bacillus thermoamylovorans] (BAH70300.1)	55 in 387 aa
230	137	-	Cobalamin synthesis protein P47K	hypothetical protein RBTH_07780 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00738433.1)	84 in 137 aa

231	103	-	Cobalamin synthesis protein P47K	hypothetical protein BTF1_32846 [Bacillus thuringiensis HD-789] (YP_006614135.1)	95 in 103 aa
232	132	-	group-specific protein	hypothetical protein pBt116 [Bacillus thuringiensis serovar israelensis] (YP_001573839.1)	93 in 132 aa
233	54	-	hypothetical protein	hypothetical protein BTF1_32336 [Bacillus thuringiensis HD-789] (YP_006614039.1)	98 in 54 aa
234	85	+	pseudogene		
235	562	-	Cry39ORF2	Cry39ORF2 protein [Bacillus thuringiensis serovar aizawai] (BAB72017.1)	83 in 558 aa
236	659	-	Cry56Ba1	Cry56Aa-like protein [Bacillus thuringiensis] (ADK38584.1)	57 in 671 aa
237	184	+	resolvase	resolvase [Bacillus cereus E33L] (YP_245576.1)	98 in 184 aa
238	467	+	Transposase for transposon Tn552	Transposase for transposon Tn552 [Bacillus cereus AH676] (ZP_04194794.1)	96 in 460 aa
239	81	+	pseudogene		
240	593	-	Fibronectin type III domain protein	hypothetical protein IIO_02806 [Bacillus cereus VD115] (ZP_17593314.1)	96 in 593 aa
241	199	-	transposon Tn552 DNA-invertase bin3	transposon Tn552 DNA-invertase bin3 [Bacillus mycoides Rock3-17] (ZP_04160431.1)	96 in 199 aa
242	982	+	Tn4652 transposase	Tn4652, transposase [Bacillus cereus Q1] (YP_002533333.1)	96 in 982 aa

LITERATURE CITED

Agaisse, H. and Lereclus, D. 1996. STAB-SD: a Shine-Dalgarno sequence in the 5' untranslated region is a determinant of mRNA stability. Mol. Microbiol. 20, 633-643.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-402.

Andrews, R.E.Jr., Bibilos, M.M. and Bulla, L.A.Jr. 1985. Protease activation of entomocidal protoxin of *Bacillus thuringiensis* subsp. *kurstaki*. Appl. Environ. Microbiol. 50, 737–732.

Andrup, L., Jensen, G.B., Wilcks, A., Smidt, L., Hoflack, L. and Mahillon, J. 2003. The patchwork nature of rolling-circle plasmids: comparison of six plasmids from two distinct *Bacillus thuringiensis* serotypes. Plasmid. 49, 205-232.

Armengol, G., Escobar, M.C., Maldonado, M.E. and Orduz, S. 2007. Diversity of Colombian strains of *Bacillus thuringiensis* with insecticidal activity against dipteran and lepidopteran insects. J. Appl. Microbiol. 102, 1, 77–88.

Aronson, A.I. 1993. The two faces of *Bacillus thuringiensis*: insecticidal proteins and post-exponential survival. Mol. Microbiol. 7, 489-496.

Aronson, A. I., Beckman, W. and Dunn. P. 1986. *Bacillus thuringiensis* and related insect pathogens. Microbiol Rev 50, 1-24.

Ash, C., Farrow, J.A., Dorsch, M., Stackebrandt, E. and Collins, M.D. 1991. Comparative analysis of *Bacillus anthracis, Bacillus cereus,* and related species on the basis of reverse transcriptase sequencing of 16S rRNA. Int. J. Syst. Bacteriol. 41, 343-346.

Barboza-Corona, J.E., Park, H.W., Bideshi, D.K. and Federici, B.A. 2012. The 60-kilodalton protein encoded by *orf2* in the *cry19A* operon of *Bacillus thuringiensis* subsp. *jegathesan* functions like a C-terminal crystallization domain. Appl. Environ. Microbiol. 78, 2005–2012.

Baum, J.A., Coyl., D.M., Gilbert., M.P., Jany., C.S. and Gawron-Burke, C. 1990. Novel cloning vectors for *Bacillus thuringiensis*. Appl. Environ. Microbiol. 56, 3420-3428.

Baum, J.A. and Gilbert., P. 1991. Characterization and comparative analysis of replication origins from three large *Bacillus thuringiensis* plasmids. J. Bacteriol. 173, 5280-5289.

Baum, J.A. and Malvar, T. 1995. Regulation of insecticidal crystal protein production in *Bacillus thuringiensis*. Mol. Microlbiol. 18, 1-12.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Wheeler, D.L. 2008. GenBank. Nucleic Acids Research. 36, 25-30.

Berry, C., Neil, S.O., Ben-Dov, E., Jones, A.F., Murphy, l., Quail, M.A., Holden, M.T.G., Harris, D., Zaritsky, A., and Parkhill, J. 2002. Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp. *israelensis*. Appl. Environ. Microbiol. 68, 5082–5095.

Boonserm, P., Davis, P., Ellar, D.J. and Li, J. 2005. Crystal structure of the mosquito -larvicidal toxin Cry4Ba and its biological implications. J. Mol. Biol. 348, 363–382.

Brantl, S., Behnke, D. and Alonso, J.C. 1990. Molecular analysis of the replication region of the conjugative *Streptococcus agalactiae* plasmid pIP501 in *Bacillus subtilis.* Comparison with plasmids pAMβ1 and pSM19035. Nucleic Acids Res. 18, 4783-4790.

Bruand, C., Lechatelier, E., Ehrlich, S.D. and Jannière, L. 1993. A fourth class of theta-replicating plasmids: The pAMβ1 family from Gram positive bacteria. Proc. Natl. Acad. Sci. USA. 90, 11668-11672.

Carroll, J., Li, J. and Ellar, D.J. 1989. Proteolytic processing of coleopteran specific delta-endotoxin produced by *Bacillus thuringiensis* var. *tenebrionis*. J. Biochem. 261, 99–105.

Carozzi, N.B., Kramer, V.C., Warren, G.W., Evola, S. and Koziel, M.G. 1991. Prediction of insecticidal activity of *Bacillus thuringiensis* strains by polymerase chain reaction product profiles. Appl. Environ. Microbiol. 57, 3057–3061.

Carlson, C.R., and Kolstø, A.B. 1993. A complete physical map of a *Bacillus thuringiensis* chromosome. J. Bacteriol. 175, 1053–1060.

Carlson, C.R., Johansen, T., Lecadet, M.M. and Kolstø, A.B. 1996. Genomic organization of the entomopathogenic bacterium *Bacillus thuringiensis* subsp. *berliner* 1715. Microbiol. 142, 1625–1634.

Carlson, C.R., Caugant, D.A. and Kolstø, A.B. 1994. Genotypic diversity among *Bacillus cereus* and *Bacillus thuringiensis* strains. Appl. Environ. Microbiol. 60, 1719–1725.

Chao, L., B. Qiyu, S. Fuping, S. Ming, H. Dafang, L. Guiming, and Y. Ziniu. 2007. Complete nucleotide sequence of pBMB67, a 67-kb plasmid from *Bacillus thuringiensis* strain YBT-1520. Plasmid. 57, 44–54.

Chattoraj, D.K. 2000. Control of plasmid DNA replication by iterons: no longer paradoxical. Mol. Microbiol. 37, 467-476.

Chen, X.J., Curtiss, A., Alcantara, E. and Dean, D.H. 1995. Mutations in domain I of *Bacillus thuringiensis* δ-endotoxin CryIAb reduced the irreversible binding of toxin to *Manduca sexta* brush border membrane vesicles. J. Biol. Chem. 270, 6412–6419.

Chen, X.J., Lee, M.K. and Dean, D.H. 1993. Site-directed mutations in a highly conserved region of *Bacillus thuringiensis* δ-endotoxin affect inhibition of

short circuit current across *Bombyx mori* midguts. Proc. Natl. Acad. Sci. USA. 90, 9041–9045.

Choi, J.Y., Kim, Y. S., Wang, Y., Tao, X.Y., Liu, Q., Roh, J.Y., Woo, S.D., Jin, B.R. and Je, Y.H. 2012. Fast and efficient generation of recombinant baculoviruses by in vitro transposition. Appl. Microbiol. Biotechnol. 96, 1353-60.

Choma, C.J. and Kaplan, H. 1991. The toxic moiety of the *Bacillus thuringiensis* protein undergoes a conformational change upon activation. Biochem. Biophys. Res. Commun. 179, 1934–1938.

Choma, C.T. and Kaplan, H. 1990. Folding and unfolding of the protoxin from *Bacillus thuringiensis*: Evidence that the toxic moiety is present in an active conformation. Biochemistry. 29, 10971–10977.

Coole, C.M. 1995. *Bacillus thuringiensis*: ecology, the significance of natural genetic modification, and regulation. World J. Microbiol. Biotechnol. 11, 471–477.

Crickmore, N., Zeigler, D.R., Schnepf, E., Van Rie, J., Lereclus, D., Baum, J., Bravo, A. and Dean, D.H. 2010. *Bacillus thuringiensis* toxin nomenclature. http://www.lifesci.sussex.ac.uk/Home/Neil_Crickmore/Bt/.

Crickmore, N., Zeigler, D.R., Schnepf, E., Van Rie, J., Lereclus, D., Baum, J., Bravo, A. and Dean, D.H. 2011. *Bacillus thuringiensis* toxin nomenclature. http://www.lifesci.sussex.ac.uk/Home/Neil_Crickmore/Bt/.

Dai, S.M. and Gill, S.S. 1993. In vitro and in vivo proteolysis of the *Bacillus thuringiensis* subsp. *israelensis* Cry IVD proteins by *Culex quinquefaciatus* larval midgut proteases. Insect Biochem. Mol. Biol. 21, 273–283.

De Barjac, H., Larget, I. and Killick-Kendrick, R. 1981. Toxicity of *Bacillus thuringiensis* var. *israelensis* serotype H14, for larvae of sandflies, the leishmaniasis vectors. Bull. Soc. Pathol. Exot. Filiales. 74, 485-489.

Dean, A.J. and Bazylinski, D.A. 1999. Genome analysis of several marine, magnetotactic bacterial strains by pulsed-field gel electrophoresis. Curr. Microbiol. 39, 219–225.

de Maagd, R.A., Bravo, A., Berry, C., Crickmore, N. and Schnepf, H.E. 2003. Structure, diversity and evolution of protein toxins from spore-forming entomopathogenic bacteria. Annu. Rev. Genet. 37, 409–433.

de Maagd, R.A., Bravo, A. and Crickmore, N. 2001. How *Bacillus thuringiensis* has evolved specific toxins to colonize the insect world. Trends Genet. 17, 193–199.

del Solar, G., Giraldo., R., Ruiz-Echevarria., M.J., Espinosa., M. and Diaz-Orejas, R. 1998. Replication and control of circular bacterial plasmids. Microbiol. Mol. Biol. Rev. 434-464.

Delcher, A. L., Harmon, D., Kasif, S., White, O. and Salzberg, S.L. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res.

27, 4636-41.

Dervyn, E., Poncet, S., Klier, A. and Rapoport, G. 1995. Transcriptional regulation of the cryIVD gene operon from *Bacillus thuringiensis* subsp. *israelensis*. J. Bacteriol. 177, 2283-91.

Ejiofor, A.O. and Johnson, T. 2002. Physiological and molecular detection of crystalliferous *Bacillus thuringiensis* strains from habitats in the South Central United States. J. Ind. Microbiol. Biotechnol. 28, 5, 284–290.

Federici, B.A. 1999. *Bacillus thuringiensis* In: Handbook of biological control. Bellows, T.S., Gordh, G., Fisher T.W. Eds. San Diego. Academic Press. 517-548.

Federici, B.A., Park, H.W. and Bideshi, D.K. 2010. Overview of the basic biology of *Bacillus thuringiensis* with emphasis on genetic engineering of bacterial larvicides for mosquito control. Open Toxinol. J. 3, 83–100.

Gamel, P.H. and Piot, J.C. 1992. Characterization and properties of a novel plasmid vector for *Bacillus thuringiensis* displaying compatibility with host plasmids. Gene. 120, 17-26.

Galitsky, N., Cody, V., Wojtczak, A., Ghosh, D., Luft, J.R., Pangborn, W. and English, L. 2001. Structure of the insecticidal bacterial delta-endotoxin Cry3Bb1 of *Bacillus thuringiensis*. Acta. Crystallogr. D. 57, 1099–1101.

Gazit, E., La Rocca, P., Sansom, M.S.P. and Shai, Y. 1998. The structure and organization within the membrane of the helices composing the pore-forming domain of *Bacillus thuringiensis* delta-endotoxin are consistent with an "umbrella-like" structure of the pore. Proc. Natl. Acad. Sci. USA. 95, 12289-12294.

Gazit, E. and Shai, Y. 1995. The Assembly and Organization of the Alpha-5 and Alpha-7 helices from the pore-forming domain of *Bacillus thuringiensis* delta-endotoxin relevance to a functional model. Journal of Biological Chemistry. 270, 2571-2578.

Gill, S.S., Cowles, E.A. and Pietrantonio, P.V. 1992. The mode of action of *Bacillus thuringiensis* endotoxins. Annu. Rev. Entomol. 37, 615–636.

González, J.M.Jr., and Carlton., B.C. 1984. A large transmissible plasmid is required for crystal toxin production in *Bacillus thuringiensis* variety *israelensis*. Plasmid. 11, 28–38.

González, J.M.Jr. and Carlton, B.C. 1980. Patterns of plasmid DNA in crystalliferous and acrystalliferous strains of *Bacillus thuringiensis*. Plasmid. 3, 92-98.

González. J.M.Jr., Dulmage, H.T. and Carlton, B.C. 1981. Correlation between specific plasmids and delta-endotoxin production in *Bacillus thuringiensis*. Plasmid. 5, 352-365.

Grochulski, P., Masson, L., Borisova, S., Pusztai-Carey, M., Schwartz, J.L.,

Brousseau, R. and Cyqler, M. 1995. *Bacillus thuringiensis* CryIA(a) insecticidal toxin: crystal structure and channel formation. J. Mol. Biol. 254, 447–464.

Guz, K., Kucinksa, J., Lonc, E. and Dorsoszkiewicz, W. 2005. Differentiated pattern of inclusions of newly isolated *Bacillus thuringiensis* strains from Silesia in Poland. Pol. J. Microbiol. 4, 263–269.

Guereca, L. and Bravo, A. 1999. The oligomeric state of *Bacillus thuringiensis* Cry toxins in solution. Biochim. Biophys. Acta. 1429, 342–350.

Heo, Y.M., Roh, J.Y., Kim, S.B., Yi, J.W., Kim, K.K., Oh, B.H. and Oh, H.T., 2012. Evaluation of the sigmoid notch involvement in the intra-articular distal radius fractures: the efficacy of computed tomography compared with plain X-ray. Clin. Orthop. Surg. 4, 83-90.

Helgason, E., Okstad, O.A., Caugant, D.A., Johansen, H.A., Fouet, A., Mock, M., Hegna, I. and Kolsto, A.B. 2000. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*-one species on the basis of genetic evidence. Appl. Environ. Microbiol. 66, 2627-2630.

Helgason, E., Caugant, D., Lecadet, M., Chen, Y., Mahillon, J., Lovgren, A., Hegna, I., Kvaløy, K. and Kolstø, A. 1998. Genetic diversity of *Bacillus cereus Bacillus thuringiensis* isolates from natural sources. Curr. Microbiol. 37, 80-87.

Hoffmaster, A.R., Ravel, J., Rasko, D.A., Chapman, G.D., Chute, M.D., Marston, C.K., De, B.K., Sacchi, C.T., Fitzgerald, C., Mayer, L.W., Maiden, M.C.J., Priest, F.G., Barker, M., Jiang, L., Cer, R.Z., Rilstone, J., Peterson, S.N., Weyant, R.S., Galloway, D.R., Read, T.D., Popovic, T. and Fraser, C.M. 2004. Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. Proc. Natl. Acad. Sci. USA. 101, 8449-8454.

Höfte, H., and Whiteley, H.R. 1989. Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiol. Rev. 53, 242-255.

Hoflack, L., Seurinck, J. and Mahillon, J. 1997. Nucleotide sequence and characterization of the cryptic *Bacillus thuringiensis* plasmid pGI3 reveal a new family of rolling circle replicons. J. Bacteriol. 179, 5000–5008.

Honee, G., Convents, D., Van Rie, J., Jansens, S., Peferoen, M. and Visser, B. 1991. The C-terminal domain of the toxic fragment of a *Bacillus thuringiensis* crystal protein determines receptor binding. Mol. Microbiol. 5, 2799-2806.

Honee, G. and Visser, B. 1993. The mode of action of *Bacillus thuringiensis* crystal proteins. Entomol. Exp. Appl. 69, 145–155.

Hu, X., Hansen, B.M., Yuan, Z., Johansen, J.E., Eilenberg, J., Hendriksen, N.B., Smidt, L. and Jensen, G.B. 2005. Transfer and expression of the mosquitocidal plasmid pBtoxis in *Bacillus cereus* group strains. FEMS.

Microbiol. Lett. 245, 239-247.

Huang, J., Guo, S., Mahillon, J., Auwera, G., Wang, Li., Han, D., Yu, Z.N. and Sun, M. 2006. Molecular characterization of a DNA fragment harboring the replicon of pBMB165 from *Bacillus thuringiensis* subsp. *tenebrionis.* BMC. Genom. 7, 260.

Hwang, S. H., Saitoh, H., Mizuki, E., Higuchi, K. and Ohba, M. 1998. A novel class of mosquitocidal delta-endotoxin, Cry19B, encoded by a *Bacillus thuringiensis* serovar *higo* gene. System. Appl. Microbiol. 21, 179-184.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W. and Hauser, L. J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. Bmc. Bioinformatics. 11.

Ibarra, J.E. and Federici, B.A. 1986. Isolation of a relatively nontoxic 65-kilodalton protein inclusion from the parasporal body of *Bacillus thuringiensis* subsp. *israelensis*. J. Bacteriol. 165, 527-533.

Ishii, T. and Ohba, M. 1993. Characterization of mosquito-specific *Bacillus thuringiensis* strains coisolated from a soil population. Syst. Appl. Microbiol. 16, 494-499.

Ito, T., Ikeya, T., Sahara, K., Bando, H. and Asano, S. 2006. Cloning and expression of two crystal protein genes, *cry30Ba1* and c*ry44Aa1*, obtained from a highly mosquitocidal strain, *Bacillus thuringiensis* subsp. *entomocidus* INA288. Appl. Environ. Microbiol. 72, 5673–5676.

Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., Bhattacharyya, A., Reznik, G., Mikhailova, N., Lapidus, A. *et al.* 2003. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. Nature, 423, 87–91.

Jensen, G.B., Wilcks, A., Petersen, S.S., Damgaard, J.J., Baum, A. and Andrup, L. 1995. The genetic basis of the aggregation system in B*acillus thuringiensis* subsp. *israelensis* is located on the large conjugative plasmid pX016. J. Bacteriol. 177, 2914–2917.

Jensen, G.B., Hansen, B.M., Eilenberg, J. and Mahillon, J. 2003. The hidden lifestyles of *Bacillus cereus* and relatives. Environm. Microbiol. 5, 631–640.

Khan, S.A. 1997. Rolling-circle replication of bacterial plasmids. Microbiol. Mol. Biol. Rev. 61, 442-455.

Koo, B.C., Kwon, M.S., Roh, J.Y., Kim, M., Kim, J.H. and Kim, T. 2012. Quantitative analysis of tetracycline-inducible expression of the green fluorescent protein gene in transgenic chickens. J. Reprod. Dev. 58, 672-677.

Kronstad, J.W., Schnepf, H.E. and Whiteley, H.R. 1983. Diversity of locations for *Bacillus thuringiensis* crystal protein genes. J. Bacteriol. 154, 419-428.

Kumar, N.S. and Venkateswerlu, G. 1997. Involvement of an endogenous metalloprotease in the activation of protoxin in *Bacillus thuringiensis* subsp.

kurstaki. Biochem. Molec. Biol. Intl. 42, 901-908.

Kumar, N.S. and Venkateswerlu, G. 1998a. Endogenous proteaseactivated 66-kDa toxin from *Bacillus thuringiensis* subsp. *kurstaki* active against *Spodoptera littoralis*. FEMS. Microbiol. Lett. 159, 113-120.

Kumar, N.S. and Venkateswerlu, G. 1998b. Intracellular proteases in sporulated *Bacillus thuringiensis* subsp. *kurstaki* and their role in protoxin activation. FEMS. Microbiol. Lett.166, 377-382.

Lecadet, M.M., Frachon, E., Dumanoir, V.C., Ripouteau, H., Hamon, S., Laurent, P. and Thiery, I. 1999. Updating the H-antigen classification of *Bacillus thuringiensis*. J. Appl. Microbiol. 86, 660-672.

Lee, H.K. and Gill, S.S. 1997. Molecular cloning and characterization of a novel mosquitocidal protein gene from *Bacillus thuringiensis* subsp. *fukuokaensis*. Appl. Environ. Microbiol. 63, 4664-4670.

Lee, I.H., Je, Y.H., Chang, J.H., Roh, J.Y., Oh, H.W., Lee, S.G., Shin, S.C. and Boo, K.S. 2001. Isolation and characterization of a *Bacillus thuringiensis* ssp. *kurstaki* strain toxic to *Spodoptera exigua* and *Culex pipiens*. Current. Microbiol. 43, 284-287.

Lenane, I.J., Bagnall, N.H., Josh, P.F., Pearson, R.D., Akhurst, R.J. and Kotze, A.C. 2008. A pair of adjacent genes, *cry5Ad* and *orf2-5Ad*, encode the typical N- and C-terminal regions of a Cry5Aδ-endotoxin as two separate proteins in *Bacillus thuringiensis* strain L366. FEMS. Microbiol. Lett. 278, 115–120.

Lereclus, D., Arantes, Q., Chaufaux, J., and Lecadet, M. 1989. Transformation and expression a cloned delta-endotoxin gene in *Bacillus thuringiensis*. FEMS. Microbiol. Lett. 51, 211–217.

Li, M.S., Je, Y.H., Lee, I.H., Chang, J.H., Roh, J.Y., Kim, H.S., Oh, H.W. and Boo, K.S. 2002. Isolation and characterization of a strain of *Bacillus thuringiensis* ssp. *kurstaki* containing a new delta-endotoxin gene. Curr. Microbiol. 45, 299-302.

Li, Z., Hiasa, H. and DiGate, R. 2005. *Bacillus cereus* DNA topoisomerase I and III alpha: purification, characterization and complementation of *Escherichia coli* TopoIII activity. Nucleic Acids Res. 33, 5415-5425.

Lukashin, A.V. and Borodovsky, M. 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26, 1107-1115.

Mahillon, J., Rezsohazy, R., Hallet, B. and Delcour, J. 1994. IS231 and other *Bacillus thuringiensis* transposable elements: a review. Genetica. 93,13-26.

Manasherob, R., Zaritsky, A., Ben-Dov, E., Saxena, D., Barak, Z. and Einav, M. 2001. Effect of accessory proteins P19 and P20 on cytolytic activity of Cyt1Aa from *Bacillus thuringiensis* subsp. *israelensis* in *Escherichia coli*. Curr. Microbiol. 43, 355-364.

Martin, P.A. and Travers, R.S. 1989. Worldwide abundance and distribution of

Bacillus thuringiensis isolates. Appl. Environ. Microbiol. 55, 2437–2442.

Nickerson, K.W. and Bulla, L.A.Jr. 1974. Physiology of sporeforming bacteria associated with insects: minimal nutritional requirements for growth, sporulation, and parasporal crystal formation of *Bacillus thuringiensis*. Appl. Microbiol. 28, 124-128.

Nishimoto, T., Yoshisue, H., Ihara, K., Sakai, H. and Komano, T. 1994. Functional analysis of block 5, one of the highly conserved amino acid sequences in the 130-kDa CryIVA protein produced by *Bacillus thuringiensis* subsp. *israelensis*. FEBS. Lett. 348, 249-254.

Noguera, P.A. and Ibarra, J.E. 2010. Detection of new *cry* genes of *Bacillus thuringiensis* by use of a novel PCR primer system. Appl. Environ. Microbiol. 76, 6150–6155.

Ohba, M. and Aizawa, K. 1978. Serological identification of *Bacillus thuringiensis* and related bacteria isolated in Japan. J. Invertebr. Pathol. 32, 303-309.

Ohba, M. and Aizawa, K. 1989. New flagellar (H) antigenic subfactors in *Bacillus thuringiensis* H serotype 3 with description of two new subspecies, *Bacillus thuringiensis* subsp. *sumiyoshiensis* (H serotype 3a:3d) and *Bacillus thuringiensis* subsp. *fukuokaensis* (H serotype 3a:3d:3e). J. Invertebr. Pathol. 54, 208-212.

Ohba, M., Mizuki, E. and Uemori, A. 2009. Parasporin, a new anticancer protein group from Bacillus thuringiensis. Anticancer Res., 29, 427–433.

Ohgushi, A., Saitoh, H., Wasano, N., Uemori, A. and Ohba, M. 2005. Cloning and characterization of two novel genes, *cry24B* and *s1orf2*, from a mosquitocidal strain of *Bacillus thuringiensis* serovar *sotto*. Curr. Microbiol. 51, 131–136.

Park, H.W., Ge, B., Bauer, L.S. and Federici, B.A. 1998. Optimization of Cry3A yields in *Bacillus thuringiensis* by use of sporulation-dependent promoter in combination with the STAB-SD mRNA sequence. Appl. Environ. Microbiol. 64, 3932–3938.

Porcar, M. and Juárez-Pérez, V.M. 2003. PCR-based identification of *Bacillus thuringiensis* pesticidal crystal genes. FEMS. Microbiol. Rev. 26, 5, 419–432.

Rasko, D. A., Altherr, M. R., Han, C. S. and Ravel, J. 2005. Genomics of the *Bacillus cereus* group of organisms. FEMS. Microbiol. Rev. 29, 303-329.

Read, T.D., Peterson, S.N., Tourasse, N., Baillie, L.W., Paulsen,I .T., Nelson ,K.E., Tettelin, H., Fouts, D.E., Eisen, J.A., Gill, S.R. *et al.* 2003. The genome sequence of *Bacillus anthracis Ames* and comparison to closely related bacteria. Nature, 423, 81–86.

Reyes-Ramirez, A. and Ibarra, J.E. 2008. Plasmid patterns of *Bacillus thuringiensis* type strains. Appl. Environ. Microbiol. 74, 125-129.

Roh, J.Y., Choi, J.Y., Jin, B.R. and Je, Y.H. 2007. *Bacillus thuringiensis* as a specific, safe, and effective tool for insect pest control. J. Microbiol.

Biotechnol. 17, 547–559.

Roh, J.Y., Liu, Q., Lee, D. W., Tao, X., Wang, Y., Shim, H.J., Choi, J.Y., Seo, J.B., Ohba, M., Mizuki, E. and Je, Y.H. 2009. *Bacillus thuringiensis* serovar *mogi* (flagellar serotype 3a3b3d), a novel serogroup with a mosquitocidal activity. J. Invertebr. Pathol. 102, 266-268.

Roh, J.Y., Park, H.W., Jin, B.R., Kim, H.S., Yu, Y.M. and Kang, S.K. 1996. Characterization of novel non-toxic *Bacillus thuringiensis* isolates from Korea. Lett. In. Appl. Microbiol. 23, 249-252.

Rosso, M. and Delécuse, A. 1997. Contribution of the 65-kilodalton protein encoded by the cloned gene *cry19A* to the mosquitocidal activity of *Bacillus thuringiensis* subsp. *jegathesan*. Appl. Environ. Microbiol. 63, 4449–4455.

Russell, R.M., Robertson, J.L. and Sauvin, S.E. 1977. POLO: A new computer program for probit analysis. Bull. Entomol. Soc. Am. 23, 209-213.

Rukmini, V., Reddy, C.Y. and Venkateswerlu, G. 2000. *Bacillus thuringiensis* crystal δ-endotoxin: role of proteases in the conversion of protoxin to toxin. Biochim. 2, 109–116.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A. and Barrell, B. 2000. Artemis: sequence visualization and annotation. Bioinfor. 16, 944-945.

Saitoh, H., Hwang, S.H., Park, Y.S., Higuchi, K., Mizuki, E. and Ohba, M. 2000. Cloning and characterization of a *Bacillus thuringiensis* serovar *higo* gene encoding a novel class of the delta-endotoxin protein, Cry27Aa, specifically active on the *Anopheles* mosquito. System. Appl. Microbiol. 23, 25-30.

Salehi, G.R., Seifinejad, A., Saeedizadeh, A., Nazarian, A., Yousefloo, M., Soheilivand, S., Mousivand, M., Jahangiri, R., Yazdani, M., Maali, A.R. and Akbari, S. 2008. Molecular detection of nematicidal crystalliferous *Bacillus thuringiensis* strains of Iran and evaluation of their toxicity on free living and plant parasitic nematodes. Can. J Microbiol. 10, 812–822.

Schneph, H.E., Whiteley, H.R. 1981. Cloning and expression of the *Bacillus thuringiensis* crystal protein gene in *Escherichia col*i. Proc. Natl. Acad. Sci. USA 78, 2893-2897.

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D.R. and Dean, D.H. 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol Mol Biol Rev. 62, 775-806.

Sezen, K., Kati, H., Muratoglu, H. and Demirbag, Z. 2010. Characterization and toxicity of *Bacillus thuringiensis* strains from hazelnut pests and fields. Pest Manag. Sci. 66, 543–548.

Sutanto, Y., Shoemaker, N.B., Gardner, J.F. and Salyers, A.A. 2002. Characterization of Exc, a novel protein required for the excision of Bacteroides conjugative transposon. Mol. Microbiol., 46, 1239–1246.

Swinfield, T., Oultram, J.D., Thompson, D.E., Brehm, J.K. and Minton, N.P. 1990. Physical characterization of the replication region of the *Streptococcus faecalis* plasmid pAMβ1. Gene. 87, 79-90.

Tang, M., Bieshi, D.K., Park, H.W. and Federici, B.A. 2006. Minireplicon from pBtoxis of *Bacillus thuringiensis* subsp. *isrelensis*. Appl. Environ. Microbiol. 72, 6984-6954.

Thomas, W.E. and Ellar, D.J. 1983. *Bacillus thuringiensis* var *israelensis* crystal delta-endotoxin: effects on insect and mammalian cells in vitro and in vivo. J. Cell Sci. 60, 181-197.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Thorne, L., Garduno, F., Thompson, T., Decker, D., Zounes, M.A., Wild, M., Walfield, A.M. and Pollock, T.J. 1986. Structural similarity between the lepidoptera- and diptera-specific insecticidal endotoxin genes of *Bacillus thuringiensis* subsp. "*kurstaki*" and "*israelensis*". J. Bacteriol. 166, 801–811.

Tinsley, E., Naqvi, A., Bourgogne, A., Koehler, T.M. and Khan, S.A. 2004. Isolation of a minireplicon of the virulence plasmid pXO2 of *Bacillus anthracis* and characterization of the plasmid-encoded RepS replication protein. J. Bacteriol. 186, 2717-2723.

Tinsley, E. and Khan, S.A. 2006. A novel FtsZ-like protein is involved in replication of the anthrax toxin-encoding pXO1 plasmid in *Bacillus anthracis*. J. Bacteriol. 188, 2829-2835.

Vaughan, S., Wickstead, B., Gull, K. and Addinall, S.G. 2004. Molecular evolution of FtsZ protein sequences encoded within the genomes of Archaea, Bacteria, and Eukaryota. J. Mol. Evol. 58, 19-39.

Vontersch, M.A., Slatin, S.L., Kulesza, C.A. and Eglish, G.H. 1994. Membrane permeabilising activities of *Bacillus thuringiensis* Cry IIIB2 and Cry IIIB2 domain I peptide. Appl Environ Microbiol. 60, 3711–3717.

Wilcks, A., Smidt., L., Okstad., O. A., Kolsto., A. B., Mahillon., J. and Andrup, L. 1999. Replication mechanism and sequence analysis of the replicon of pAW63, a conjugative plasmid from *Bacillus thuringiensis*. J. Bacteriol. 181, 3193 -3200.

Wu, D. and Federici, B.A., 1993. A 20-kilodalton protein preserves cell viability and promotes cyta crystal-formation during sporulation in *Bacillus thuringiensis*. J. Bacteriol. 175, 5276-5280.

Xu, Y.N., Uhm, S.J., Koo, B.C., Kwon, M.S., Roh, J.Y., Yang, J.S., Choi, H.Y., Heo, Y.T., Cui, X.S., Yoon, J.H., Ko, D.H., Kim, T. and Kim, N.H. 2013. Production of transgenic Korean native cattle expressing enhanced green

fluorescent protein using a FIV-based lentiviral vector injected into MII oocytes. J. Genet. Genomics. 40, 37-43.

Yamagiwa, M., Esaki, M., Otake, K., Inagaki, M., Komano, T., Amachi, T. and Sakai, H. 1999. Activation process of dipteran-specific insecticidal protein produced by *Bacillus thuringiensis* subsp. *israelensis*. Appl. Environ. Microbiol. 65, 3464-3469.

Yamamoto, T. and Powell, G.K. 1993. *Bacillus thuringiensis* crystal proteins: recent advances in understanding its insecticidal activity. In: Kim, L. (Ed.), Advanced Engineered Pesticides. Marcel Dekker Inc., New York, pp. 3–42.

Zhao, X., Zhang, Z., Yan, J. and Yu, J. 2007. GC content variability of eubacteria is governed by the pol III alpha subunit. Biochem. Biophys. Res. Commun. 356, 20-25.

Zhao, X.Q., Hu, J.F. and Yu, J. 2006. Comparative analysis of eubacterial DNA polymerase III alpha subunits. Genomics Proteomics Bioinfomatics. 4, 203-211.

모기 살충성 strain, *Bacillus thuringiensis* subspecies *mogi* 의 분자 생물학적 특성 연구

서울대학교

농생명공학부 곤충학전공

유금

초 록

국내 문경지역에서 수집한 낙엽으로부터 분리한 *Bacillus thuringiensis* subsp. *mogi* 균주는 plasmid 상에 곤충병원성과 관련된 다양한 유전자를 가지고 있다. 따라서 본 연구에서는 *B. thuringiensis* subsp. *mogi* 균주의 생물학적 특성을 구명하고, 전체 genome 염기서열 및 유전자 구조를 분석하며, plasmid 상에 위치한 새로운 cry 유전자의 발현에 대한 분자생물학적 특성을 조사하고자 하였다.

55 개 type-strain 의 편모 항혈청을 이용하여 편모항원성을 검정한 결과 *B. thuringiensis* subsp. *mogi* 균주는 3a3b3d의 새로운 serogroup인 것으로 나타났으며, 이 결과를 바탕으로 serovar *mogi* 로 명명하였다. 이러한 *B. thuringiensis* subsp. *mogi* 균주는 *Culex pipiens molestus* 와 *Culex pipiens pallens* 등 파리목 유충에 대해서는 살충활성을 보였지만 나비목 유충에 대해서는 살충활성을 보이지 않았다. 또한, *B. thuringiensis* subsp. *mogi* 균주는 3 개의 난형의 parasporal crystal 이 하나의 envelope 에 둘러싸인 형태의 inclusion body 를 생성하였으며, SDS-PAGE 를 수행한 결과 이들 parasporal crystal 은 30~75 kDa 정도의 분자량을 가진 여러개의 단백질로 이루어져 있는 것으로

나타났다. 이들 단백질 band 에 대하여 nano-LC-ESI-IT MS 분석을 수행한 결과, Cry27Aa, Cry39ORF2 및 Cry20-like 의 putative peptides 인 것으로 확인되었다. 한편, H3 serotype 에 속하는 기존의 *B. thuringiensis* 균주들은 복잡한 plasmid profile 을 보이는데 비해 *B. thuringiensis* subsp. *mogi* 균주는 30 MDa 이상의 megaplasmid 만 보유하고 있는 것으로 나타났다.

B. thuringiensis subsp. *mogi* 균주의 전체 genome 은 약 6.0 Mb 였으며, 5,652 개의 ORF 를 coding 하고 있는 circular chromosome (약 5.4 Mb)과 두개의 megaplasmid, pMOGI364 (364,564 bp) 및 pMOGI222 (222,348 bp) 등 총 세 개의 replicon 으로 이루어져 있었고, 이들 replicon 의 G+C contents 는 31.3~35.2%였다. 두 개의 megaplasmid 상에는 모두 17 개의 병원성 관련 cry 유전자가 존재하는 것으로 분석되었으며, 이중 6 개의 유전자 (*cry19Bb1*, *cry73Aa* 과 *cry4orf2* operon, *cry20Bb1, cry27Ab1, cry4Aa, cry56Ba1* 과 *cry39orf2* operon)가 기존의 *cry* 유전자에서 보고된 세 개의 domain 을 모두 가지는 것으로 보아 실질적인 살충활성을 가질 것으로 예상되었다.

앞에서 살충활성을 가질 것으로 예상된 6 개의 새로운 *cry* 유전자에 대하여 정량 PCR (qPCR) 을 수행한 결과, 이들 유전자 모두가 *B. thuringiensis* subsp. *mogi* 균주 내에서 정상적으로 transcription 이 되는 것을 확인할 수 있었다. 이들 *cry* 유전자의 발현 특성을 알아보기 위하여 자신의 promoter 의 조절 하에서 *Escherichia coli-B. thuringiensis* shuttle vector 인 pHT1K 에 cloning 하고 acrystalliferous *B. thuringiensis* Cry-B 균주에 도입한 결과, *cry20Bb1* 과 *cry56Ba1* operon 이 형질전환된 균주에서만 wild-type *B.*

thuringiensis subsp. *mogi* 균주에서보다 그 크기는 작지만 crystal 을 형성하였으며, crystal 을 형성한 경우에만 모기 유충에 살충활성을 보였다. 한편, *cry56Ba1* operon 에서 *cry39orf2* 의 역할을 알아보기 위하여 이들 유전자를 STAB-SD sequence 와 강력한 chimeric *cyt1Aa* promoter 를 가진 over-expression vector 인 p1KSD 에 cloning 하고 acrystalliferous *B. thuringiensis* Cry-B 균주에 도입하였다. 그 결과, intact 한 operon 구조뿐만 아니라 *cry39orf2* 만 발현시켜도 crystal 을 형성하는 것을 확인할 수 있었다. 이러한 결과는 Cry39ORF2 가 기존의 Cry 단백질에서 structural region 으로 보고된 C-말단의 역할을 하여 Cry59Ba1 의 crystallization 에 관여한다는 것을 암시하였다.

검 색 어 : *B.thuringiensis,* ovoidal-shaped crystals, mosquitocidal, full genome sequence, three-domain *cry* gene, over-expression

학번: 2007-23602

감사의 글

2013년의 겨울은 기다림 속에서 맞이하였습니다. 서울대학교에서 보낸 6년반의 세월동안 좋으신 분들을 많이 만날수 있었던 것은 저에게는 큰 행운이었습니다. 박사학위 논문이 끝나는 이 시점에서, 저는 저에게 도움을 주신 많은 고마운 분들께 깊은 감사의 인사를 올리고 싶습니다.

먼저, 저의 지도교수님이신 제연호 교수님께 가장 진심 어린 경의와 고마움을 전하고 싶습니다. 저는 교수님의 학생이 된 것을 매우 영광으로 생각합니다. 교수님은 일상생활과 학업에서 모두 주도면밀한 관심을 주셨는바, 교수님의 은혜는 평생 가슴에 새기도록 하겠습니다. 학문을 대하는 교수님의 태도, 넓은 흉금, 정직한 인격, 긍정적인 에너지는 제가 평생 따라 배워야 하는 본보기입니다. 언제 어디서나 교수님 생각만 하면 저의 마음속에는 더없는 자부감이 생깁니다. 저도 교수님과 같은 사람으로 거듭나도록 노력할 것입니다. 재학 동안, 제가 교수님께 끼친 폐에 대하여 매우 미안한 마음이 있으며 동시에 교수님께서 너그럽게 헤아려 주시고 이해해 주신데 대해 고마울 따름입니다.

또한 저의 눈문심사위원이신 안용준 교수님, 이승환 교수님, 국립보건연구원 노종열 박사님, 전북대학교 김재수 박사님께 감사드립니다. 교수님들은 바쁘신 와중에도 보귀한 시간을 할애하여 저의 논문을 읽어주시고 수정해 주셨으며 많은 소중한 견해를 제기하여 주셨습니다.

그리고 학과의 이준호 교수님, 이시혁 교수님, 이광범 교수님한테 고마움을 전하고 싶습니다. 여러해 동안 교수님들의 생동감있고 풍부한 강의를

감명깊게 들었으며 교수님들로 하여 서울대학교 곤충전공은 현재와 같이 나날이 발전할 수 있을 것이라 생각됩니다.

심혈을 기울여 저의 실험을 첫번째로 도와주시고 *Bacillus thuringiensis* 의 세계로 이끌어주신 노종열 박사님한테 감사의 인사를 올립니다. 박사님으로 부터 저는 하나의 완전한 실험방안의 건립방법을 배웠으며 과학연구 과정에서 즐거움을 찾는 방법을 배웠습니다.

그리고 실험실의 학우들-최재영 박사님, 김우진 박사님, 후배 박종빈, 이주현, 김송은, 이석희, 안샛별, 방영, 김종훈에게 고마움을 전하고 싶습니다. 이들은 저에게 여러 면에서 도움을 주었고 실험실에서 함께 제가 발전하는 모습을 지켜봐 주었습니다. 우리의 "행복한 대가정" 성원들이 영원히 행복하길 바랍니다.

또한 임재윤 박사님의 아낌없는 도움에 감사를 드립니다.

동시에 한국땅에 밟은 후 첫번째로 친절하게 다가왔고, 항상 오빠처럼 저를 보살펴 준 Wang Yong형에게 고마우을 전하고, Tao Xueying과는 함께 5년이라는 시간을 보냈는데 우리의 방황, 고통, 기쁨은 서로에게 제일 좋은 격려가 되었습니다. Tao Xueying와 Tao Xueying의 아이가 늘 건강하고 행복하길 바랍니다.

그리고 잊지못할 나날들을 함께 보낸 Lin Yang, Cheng Jing, Liu Yang, Gao Yaqian, Jin Yinhua , Gao Zhe, Fu zhenli, Wang zhangqian, Wang xue, Zhang Xiangmei, He Hai, Cheng Yanpin, Liu Xiaoshan, Wu Lili, Song Xiangqing , Shi Siyu , Jing Liang, 등 나의 친구들에게 감사드립니다. 친구들

은 저한테 많은 관심과 격려를 주었고 그들로인해 나의 한국 생할은 더욱 즐겁고 따뜻했습니다. 친구들에게 많은 것을 배웠으며 함께 있었던 나날들은 참으로 즐거운 시간들이었습니다.

마지막으로 저의 아버지, 어머니, 여동생, 남동생한테 감사의 인사를 올립니다. 그들은 저에게 아낌없는 사랑을 주었고 저의 학업을 위해 거대한 희생과 노력을 대가로 하였다는 것을 잘 알고 있지만 저는 아직 그 은혜를 갚지 못했습니다.

저의 가족들께, 그리고 저를 관심하여 주신 모든 여러분께 다시한번 무한한 축복을 드립니다. 여러분의 행복과 건강을 기원합니다.

A THESIS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Molecular Biological Characterization of Mosquitocidal Strain, *Bacillus thuringiensis* subspecies *mogi*

모기 살충성 strain, *Bacillus thuringiensis* subspecies *mogi* 의 분자 생물학적 특성 연구

By

Qin Liu

Department of Agricultural Biotechnology

Seoul National University

February, 2014

Molecular Biological Characterization of Mosquitocidal Strain, *Bacillus thuringiensis* subspecies *mogi*

Major in Entomology

Department of Agricultural Biotechnology, Seoul National University

Qin Liu

ABSTRACT

Bacillus thuringiensis subspecies *mogi* was isolated from fallen leaves, sampled in a forest region of the city of Mungyeong, Korea. Plasmids from this species have been implicated in pathogenicity as they carry genes responsible for a variety of entomo-pathogenic diseases. The purpose of this study was to characterize the *B. thuringiensis* subsp. *mogi* strian, determine the full genome sequence, and investigate the molecular genetics of expression of novel toxin-related *cry* genes which located on the plasmid in *B. thuringiensis* subsp. *mogi*.

As a primary study, the flagellated vegetative cells of *B. thuringiensis* subsp. *mogi* were agglutinated with the H3 reference antiserum and further agglutinated with 3b and 3d monospecific antisera but non-reactive to 3c and 3e factor sera. These results

create a new serogroup with flagellar antigenic structure of 3a3b3d, designated serovar *mogi*. *B. thuringiensis* subsp. *mogi* showed activity against dipteran larvae, *Culex pipiens molestus* and *Culex pipiens pallens* while no lepidopteran toxicity. It produced three small ovoidal-shaped parasporal crystals combined together and whose SDS-PAGE protein profile consisted of several bands ranging from 75 to 30 kDa. Through the identification of the protein by nano-LC-ESI-IT MS analysis, the putative peptides of Cry27Aa, Cry39ORF2, and Cry20-like were detected. In contrast to the complicated plasmid profiles of *B. thuringiensis* H3 serotype strains, the *B. thuringiensis* subsp. *mogi* contained only megaplasmids (> 30 MDa) on which the toxin genes were occasionally located.

Second, full genome sequence of the novel *B. thuringiensis* subsp. *mogi* strain was determined. The 6.0 Mb genome of *B. thuringiensis* subsp. *mogi* contains three replicons: a circular chromosome (5.40 Mb) encoding 5,652 predicted open reading frames (ORFs) and two mega-plasmids, pMOGI364 (364,564 bp) and pMOGI222 (222,348 bp). The G+C contents of these replicons ranged from 31.3% to 34.2% for pMOGI364 and pMOGI222, respectively, and did not significantly deviated from that of the chromosome (35.2%). There were seventeen toxin-related genes existed in these two mega-plasmids, and six of them (*cry19Bb1, cry73Aa* with *cry40orf2, cry20Bb1, cry27Ab1, cry4Aa* and *cry56Ba1* with *cry39orf2*) belonged to the group of three-domain *cry* toxins.

Finally, to investigate the role of six novel three-domain *cry* genes in crystal production of *B. thuringiensis* subsp. *mogi*, the transcription level of these toxin genes were analyzed by quantitative PCR (qPCR). The results clearly indicated that all of these *cry* genes were successfully transcribed in wild type *B. thuringiensis* subsp. *mogi* strain in different transcription time with different maximum levels. Then, these *cry* genes were cloned to the *Escherichia coli-B. thuringiensis* shuttle vector, pHT1K, under the control of its own promoter, and introduced into an acrystalliferous *B. thuringiensis* Cry$^-$B strain for further molecular characterization. Another vector p1KSD, which containing a strong chimeric *cyt1Aa* promoter combined with the STAB-SD sequence was constructed and used to over-express the *cry* genes. To determine the function of the *cry39orf2* and over-express the *cry56Ba1* in *cry56Ba1* operon, different combinations of Cry56Ba1 and Cry39ORF2 were synthesized in strain Cry$^-$B. The stable inclusion in recombinant cells suggests that Cry39ORF2 assists in synthesis and crystallization of Cry56Ba1 by functioning like the C-terminal domain characteristic of Cry protein in the 130 kDa mass range. In addition, the increased Cry56Ba1 yield under the *cyt1A-p*/STAB-SD promoter has broadened the possibility of application in other toxins.

Key words: *B.thuringiensis,* ovoidal-shaped crystals, mosquitocidal, full genome sequence, three-domain *cry* gene, over-expression

Student number: 2007-23602

TABLE OF CONTENTS

ABSTRACT----------i

TABLE OF CONTENTS ----------iv

LIST OF TABLES ----------vi

LIST OF FIGURES ----------viii

LIST OF ABBREVIATIONS----------xi

LITERATURE REVIEW----------xii

Chapter 1. Characterization of a novel serogroup *Bacillus thuringiensis* strain, subsp. *mogi,* flagellar serotype 3a3b3d

ABSTRACT ----------1

INTRODUCTION----------2

MATERIALS AND METHODS----------3

RESULTS ----------15

DISCUSSION----------25

Chapter 2. Genome sequencing strategy and sequence analysis of *Bacillus thuringiensis* subsp. *mogi*

ABSTRACT ----------28

INTRODUCTION----------29

MATERIALS AND METHODS----------32

RESULTS ----------34

DISCUSSION----------72

Chapter 3. Molecular cloning and characterization of mosquitocidal protein genes from *B. thuringiensis* subsp. *mogi*

ABSTRACT ----------74

INTRODUCTION----------75

MATERIALS AND METHODS----------77

RESULTS ----------95

DISCUSSION----------141

LITERATURE CITED----------186

ABSTRACT IN KOREAN----------195

ACKNOWLEDGEMENTS----------198

LIST OF TABLES

Table 1. Composition of culture media for a new strain of *B. thuringiensis*----------5

Table 2. H agglutination test results of *B. thuringiensis mogi* strain----------7

Table 2-1. Monospecific antisera agglutination test results of *B. thuringiensis mogi* strain----------8

Table 3. Nucleotide sequences of primers used for amplification of the specific probe in southern blot----------14

Table 4. Toxicity of *B. thuringiensis* against *Culex pipiens molestus* and *Culex pipens pallens* 4th instar larvae----------24

Table 5. Read status of *B. thuringiensis* subsp. *mogi* genome assemblies----------36

Table 6. General characteristics of chromosome and gene prediction----------37

Table 7. General characteristics of 2 plasmids from strain *B. thuringiensis* subsp. *mogi*----------37

Table 8. DNA topoisomerase III in *B. thuringiensis mogi.*----------54

Table 9. Nucleotide sequences of forward and reverse primers used for amplification of the target *cry* genes in reverse transcription-polymerase chain reaction (RT-PCR)----------80

Table 10. Nucleotide sequences of primers used for amplification of the target *cry* genes in qPCR----------82

Table 11. Nucleotide sequences of primers used for amplification of the target *cry* genes to pHT1K vector in PCR----------85

Table 12. Primers used for amplification of *cry56Ba1* and *39orf2*----------91

Table 13. Summary of toxin-related genes in pMOGI364 and pMOGI222----------98

Table 14. Amino acid sequences of putative crystal proteins of *B. thuringiensis* subsp. *mogi* strain----------99

Table 15. Toxicity of recombinant strains of *B. thuringiensis* Cry-B against *Culex pipiens molestus* and *Cluex pipiens pallens* 4th instar larvae----------133

Table 16. Toxicity of recombinant strains of *B. thuringiensis* Cry-B against *Culex pipiens molestus* and *Cluex pipiens pallens* 4th instar larvae----------139

Table S1. The CDS in pMOGI364 and their annotations----------146

Table S2. The CDS in pMOGI222 and their annotations----------170

LIST OF FIGURES

Fig. 1. Plasmid patterns from *B. thuringiensis* H3 type strains----------16

Fig. 1-1. Plasmid patterns from *B. thuringiensis* mosquitocidal type strains----------17

Fig. 2. Plasmid pattern from *B. thuringiensis* subsp. *mogi* strain
was examined in pulse field gel electrophoresis----------18

Fig. 3. Transmission electron microscopy of the parasporal crystal
of *B. thuringiensis* subsp. *mogi* strain----------21

Fig. 4. SDS-PAGE and nano-LC-ESI-IT MS analysis----------22

Fig. 5. Southern hybridization of *B. thuringiensis* subsp. *mogi*----------23

Fig. 6. GC-skew analysis of chromosome----------38

Fig. 7. Mummer analysis of *B. thuringiensis mogi* strain with the
selected *Bacillus* species----------39

Fig. 8. Phylogenetic relationship of the *B. thuringiensis mogi*
among the *Bacillus* clade----------42

Fig. 9. Circular representation of pMOGI364----------49

Fig. 10. GC-skew of the plasmid pMOGI364----------51

Fig. 11. Mummer analysis of pMOGI364 with pG9842_209----------52

Fig. 12. Graphic view of pMOGI364 megablast results----------53

Fig. 13. Previews about topoisomerase III in *B. thuringiensis mogi*----------55

Fig. 14. Sequence alignment analysis of the truncated DNA topoisomerase III
with the intact one from NCBI database----------56

Fig. 15. Circular representation of pMOGI222----------64

Fig. 16. GC-skew of the plasmid pMOGI222----------66

Fig. 17. Graphic view of pMOGI222 megablast results----------67

Fig. 18. Alignment of the predict Rep protein of pMOGI222, Rep165 protein
of plasmid pBMB165, Rep63A protein of plasmid pAW63, RepS protein
of pXO2, and RepE protein of pAMβ1 of these three plasmids----------68

Fig. 19. Alignment of the origin of replication in pMOGI222, pBMB165, pBT9727, pAW63, pXO2, pAMβ1 and p43----------70
Fig. 20. The pMOGI222 replication region----------71
Fig. 21. Comparison and clustering of *B. cereus* group plasmids based on replication----------72
Fig. 22. Cloning and expression analysis of mosquitocidal *cry* genes from *B. thuringiensis* subsp. *mogi*----------86
Fig. 23. Schematic illustration of the different constructs containing *cry56Ba1* or *cry39orf2* ----------90
Fig. 24. Upstream region and schematic illustration of constructs used to synthesize Cry proteins----------91
Fig. 25. The over-expression vector p1KSD----------93
Fig. 26. Nucleotide sequence and deduced amino acid sequence of the *cry19Bb1* gene----------100
Fig. 27 Nucleotide sequence and deduced amino acid sequence of the *cry73Aa* and *cry40orf2* gene----------103
Fig. 28. Nucleotide sequence and deduced amino acid sequence of the *cry20Bb1* gene----------107
Fig. 29. Nucleotide sequence and deduced amino acid sequence of the *cry27Ab1* gene----------110
Fig. 30. Nucleotide sequence and deduced amino acid sequence of the *cry4Aa* gene----------113
Fig. 31. Nucleotide sequence and deduced amino acid sequence of the *cry56Ba1* and *cry39orf2* gene----------116
Fig. 32. Comparison of the deduced amino acid sequence of Cry19Bb1, Cry20Bb1, Cry73Aa with other Cry proteins in five conserved blocks----------121
Fig. 33. Comparison of the deduced amino acid sequence of Cry4Aa, Cry27Ab1,

Cry56Ba1 with other Cry proteins in five conserved blocks----------123
Fig. 34. Agarose gel electrophoresis analysis of *cry* genes mRNA obtained by RT-PCR from *B. thuringiensis* subsp. *mogi* in the sporulation stage----------125
Fig. 35. Transcription-level analysis of the six *cry* genes from *B. thuringiensis* subsp. *mogi* were analyzed----------126
Fig. 36. Agarose gel electrophoresis analysis of PCR products----------129
Fig. 37. SDS-PAGE analysis of the recombinant *B. thuringiensis cry*⁻*B* strains containing different *cry* genes----------130
Fig. 38. Phase contrast micrographs of the recombinant *B. thuringiensis* Cry-B strains containing different *cry* genes----------131
Fig. 39. Transmission electron microscopy of the acrystalliferous *B. thuringiensis* Cry-B strain and the recombinant strains----------132
Fig. 40. Over-expression of *cry56Ba1* operon and functional analysis of *cry39orf2*----------137
Fig. 41. Alignment of the ORF2 amino acid sequence with C terminal regions of selected Cry proteins----------140

LIST OF ABBREVIATIONS

BHI: brain heart infusion

bp: base pair(s)

ca.: approximately

DIG: dioxigenin

DNA: deoxyribonucleic acid

dNTP: dexoxyribonucleoside triphosphate

EDTA: ethylenediaminetetra acetic acid

g: acceleration due to gravity

GYS: glucose-yeast extract salt medium

h: hour(s)

kb: kilo base pair(s)

kDa: kilo Dalton

LB: Luria-Bertani medium

min: minute(s)

M: molarity (= mol/l)

OD: optical density

PAGE: polyacrylamide gel electrophoresis

PVDF: polyvinylidene fluoride

LITERATURE REVIEW

Basic biology of *Bacillus thuringiensis*

B. thuringiensis Berliner was originally discovered in Japan over a century ago by Shigetane Ishiwata (Federici *et al*., 2010) as the cause of the sudden ("sotto") death disease of silkworms, larvae of the silkworm moth, *Bombyx mori*. Ten years later, the German bacteriologist Ernst Berliner, unaware of Ishiwata's paper, described a similar bacterium as the cause of disease in larvae of the flour moth, *Ephestia kuhniella*. The species name "*thuringiensis*" is derived from Thuringia, the German state where the diseased flour moth larvae were found.

The gram-positive bacterium *B. thuringiensis*, which can be readily isolated from a variety of environmental sources including soil, water, plant surfaces, grain dust, dead insects, and insect feces (Federici, 1999), was characterized by its ability to produce crystalline inclusions during sporulation. Its life cycle is simple. When nutrients and environmental conditions are sufficient, the spore germinates producing a vegetative cell that grows and reproduces by binary fission. Cells continue to multiply until one or more nutrients, such as sugars, amino acids, or oxygen, become insufficient for continued vegetative growth. Under these conditions, the bacterium sporulates producing a spore and parasporal body, the latter, composed primarily of one or more insecticidal proteins in the form of crystalline inclusions (Federici *et al*., 2010). These

are commonly referred to in the literature as insecticidal crystal proteins or δ-endotoxins, which can compose as much as 40% of the dry weight of a sporulated culture. These inclusions consist of proteins exhibiting a highly specific insecticidal activity (Aronson *et al*., 1986). Most crystal proteins are active against larvae of certain members of the Lepidoptera, but some show toxicity against dipteran (flies) or coleopteran (beetles) insects, or nematodes.

Insecticidal proteins in *B. thuringiensis*

There are two types of insecticidal crystal proteins in *B. thuringiensis*, Cry (for crystal) and Cyt (for cytolytic) proteins, and variations on each of these types. Individual Cry toxins have a defined spectrum of insecticidal activity, usually restricted to a few species within one particular order of insects. To date, toxins for insect species in the orders Lepidoptera (butterflies and moths), Diptera (flies and mosquitoes), Coleoptera (beetles and weevils) and Hymenoptera (wasps and bees) have been identified. The Cry proteins are classified on the basis of amino acid sequence homology, where each protoxin acquired a name consisting of the mnemonic Cry (or Cyt) and four hierarchical ranks consisting of numbers, capital letters, lower case letters and numbers (e.g. Cry25Aa1), depending on its place in a phylogenetic tree. The know Cry and Cyt proteins now fall into 32 sets including Cyt1, Cyt2 and Cry1 to Cry 67 (Crickmore *et al.*, 2010).

The structural diversity of Cry toxin

In *B. thuringiensis*, these proteins form crystals, with the most common types being composed of Cry1 proteins of about 135 kDa. These are primarily toxic to lepidopterous insects, and consist of a N-terminal half (contains the active protein) containing the toxic portion of the molecule, released after ingestion by insect midgut proteases, and a C-terminal half important to crystallization (Schnepf *et al*., 1998). In addition to 135 kDa proteins, Cry proteins of 65-70 kDa are known which correspond to the N-terminal half of the 135 kDa Cry type. Examples include Cry2A toxic to lepidopterous and dipterous insects, Cry3A (similar in mass to Cry2) toxic to coleopterous insects, and Cry11A toxic to certain dipterous insects. Phylogenetic studies indicate that all of the above Cry types evolved over millions of years from the same ancestral molecule, the diversity in host spectra being selected for when mutant strains wound up in the midguts of insect species belonging to different orders.

Five highly conserved blocks exist in the toxic core of most known Cry protoxins, which are important for their activities and specificities (Höfte and Whiteley, 1989). They are arranged in three distinct domains (I–III, from N- to C-termini). Block 1, encompassing the central helix α5 of domain I, has been implicated in pore formation, a role that might explain its highly conserved nature (Gazit *et al.*, 1998). Block 2 includes the C-terminal half of helix α6 and all of α7 of domain I, and the first β-strand of domain II. Helix α7 serves as a binding sensor to initiate the structural

rearrangement of the pore-forming domain (Gazit and Shai, 1995). Residues within block 2 are involved in formation of salt bridges, which could be considerable, in conformational changes upon binding of the toxin to receptor or for maintaining the protein in globular form (Schnepf *et al.,* 1998). Block 3 contains the last β-strand of domain II and the N-terminal segment of domain III, the latter forming the interface with domains I and II. Block 4 corresponds to the second β-strand of domain III that affects the structural integrity of the protein, oligomeric aggregation, and the appropriate function of the ion channels. The highly conserved block 5 in domain III is at the C-terminus of the activated toxin and is another major element that stabilizes the mature toxin (Nishimoto *et al.,* 1994; Yamagiwa *et al.,* 1999)

Endogenous proteases in *B. thuringiensis*

During the early sporulation phase, an increase in intracellular protease activity occurs in *B. thuringiensis* cultures. Proteases endogenous to *B. thuringiensis* have been described from the cysteine, metallo, and serine families of enzymes. Major proteases in most *B. thuringiensis* species are thermostable and many are metalloproteases, with some exceptions. The endogenous proteolytic activities in *B. thuringiensis* may hydrolyze crystal proteins. For example, a reduction in the size of *B. thuringiensis* subsp. *tenebrionis* inclusion crystal proteins (ICPs) occurred during sporulation, and proteolysis was prevented by the addition of protease inhibitors (Carroll *et al.*, 1989). ICPs from *B. thuringiensis* subsp. *kurstaki* crystals, incubated in

denaturing and reducing conditions, were hydrolyzed by metalloproteases in the crystal (Kumar and Venkateswerlu, 1997). Interestingly, the toxin produced under these conditions was highly active against the cotton leafworm, *Spodoptera littoralis*, a species insensitive to native *kurstaki* crystals or toxins generated by exogenous proteases (Kumar and Venkateswerlu, 1998a). Although the crystal contained multiple Cry proteins, the toxin was homogenous, as demonstrated by two-dimensional polyacrylamide gel electrophoresis, and lacked the first 29 amino acids of the protoxin N-terminus (Kumar and Venkateswerlu, 1998b). Mosquitocidal ICPs were also degraded in the crystal (Dai and Gill, 1993). The mosquitocidal protoxin Cry11Aa1 was partially processed from 72 to 32 – 40 kDa proteins within the crystal by endogenous *B. thuringiensis* proteases (Ibarra and Federici, 1986).

Chapter 1. Characterization of a novel serogroup *Bacillus thuringiensis* strain, subsp. *mogi*, flagellar serotype 3a3b3d

ABSTRCT

Bacillus thuringiensis strain *mogi* was isolated from fallen leaves, sampled in a forest region of the city of Mungyeong, Korea. The flagellated vegetative cells of *B. thuringiensis* strain were agglutinated with the H3 reference antiserum and further agglutinated with 3b and 3d monospecific antisera but non-reactive to 3c and 3e factor sera. These results create a new serogroup with flagellar antigenic structure of 3a3b3d, designated serovar *mogi*. The strain *mogi* showed activity against dipteran larvae, *Culex pipiens molestus* and *Culex pipiens pallens* while no lepidopteran toxicity. It produced three small ovoidal parasporal crystals combined together and whose SDS-PAGE protein profile consisted of several bands ranging from 75 to 30 kDa. Through the identification of the protein by nano-LC-ESI-IT MS analysis, the putative peptides of Cry27Aa, Cry39ORF2, and Cry20-like were detected. In contrast to the complicated plasmid profiles of *B. thuringiensis* H3 serotype strains, the *B. thuringiensis* subsp. *mogi* contained only megaplasmids (> 30 MDa) on which the toxin genes were occasionally located. The new type strain, *B. thuringiensis* subsp. *mogi* (H3a3b3d) will be a good resource for novel mosquitocidal *cry* genes.

Key words: *Bacillus thuringiensis*, novel serogroup, mosquitocidal, *cry* genes

1. INTRODUCTION

The Gram-positive and endospore-forming bacterium *Bacillus thuringiensis*, which is frequently used in industrial applications, is well known for its ability to produce crystalline parasporal inclusions that have insecticidal activity against various species. The parasporal inclusion, which may contain more than one type of insecticidal crystal protein (ICPs), is released with the spore upon lysis of the sporangium (Höfte and Whiteley, 1989; Schnepf *et al*., 1998). A number of isolates of the bacterium are commercially produced, with activity against Lepidoptera, Diptera and Coleoptera.

B. thuringiensis produces parasporal inclusions (crystals) having several unique features including insecticidal, nematocidal or anti-cancer activity (Ohba *et al.*, 2009; Roh *et al*., 2007). Numerous *B. thuringiensis* isolates have been collected worldwide and some of them have been characterized by various techniques such as biochemical test, H-serotyping, plasmid patterns and *cry* gene contents by PCR analysis (Lecadet *et al.*, 1999; Porcar and Juarez-Perez, 2003; Reyes-Ramirez and Ibarra, 2008). Among them, the classification of *B. thuringiensis* isolates by H-serotyping has been believed as an efficient way since it is based on the stable and specific characters of the flagellar (H) antigen. The H-serotyping, however, has limitations, proving unreliable as a predictor of insecticidal activity. For example, *B. thuringiensis* serovar *morrisoni* (H8a8b) is a collection of heterogeneous pathovars specifically active against Lepidoptera, Coleoptera, or Diptera, and even those with no insecticidal activities

(Park *et al*., 1998). Nevertheless, it is still of great value in discriminating between *B. thuringiensis* strains (Lecadet *et al*., 1999). By the end of 1998, up to 69 different serotypes and 13 sub-antigenic groups, giving 82 serovars, have been involved in H-serotype classification scheme.

B. thuringiensis is widely distributed and recovered from 70% of soil samples from all continents, with Asian samples being an especially rich source; *B. thuringiensis* subsp. *israelensis* and subsp. *kurstaki* are the most common types (Martin and Travers, 1989). It is always desirable to search for a better insecticide against noxious insects (Sezen *et al*., 2010). In this study, a new subserotype *B. thuringiensis* strain, occurring in the H3 serogroup, which has larvicidal activity against *Culex* mosquitoes, was isolated and characterized.

2. MATERIALS AND METHODS

2.1 Bacterial strains and growth media

The *B. thuringiensis* strain *mogi* was isolated from fallen leaves, sampled in a forest region of the city of Mungyeong, Korea, according to the method of Ohba and Aizawa (Ohba and Aizawa, 1978). Other *B. thuringiensis* type strains include in this work were kindly provides by the International Entomopathogenic *Bacillus* Center (IEBC) at the Pasteur Institute, Paris, France. *B. thuringiensis* were grown at 28ºC

with vigorous shaking in SPY medium for plasmid preparation and GYS medium for expression of crystal proteins (Kronstad *et al*., 1983; Li *et al.*, 2002; Nickerson and Bulla, 1974). The LB medium was used as a primary culture of *B. thuringiensis* and in *E.coli* culture for plasmid preparation. Media compositions were described in Table 1. Brain heart infusion (BHI) medium was used to culture competent *B. thuringiensis* cells.

Table 1. Composition of culture media for a new strain of *B. thuringiensis.*

Medium*	Component	% (g/L)
LB	Trypton	1
	Yeast extract	0.5
	Nacl	1
GYS	Glucose	0.1
	Yeast extract	0.2
	K_2HPO_4	0.05
	$(NH_4)_2SO_4$	0.2
	$MgSO_4$	0.002
	$MnSO_4$	0.005
	$CaCl_2$	0.008
SPY	$(NH_4)_2SO_4$	0.2
	K_2HPO_4	1.4
	KH_2PO_4	0.6
	$Na_3C_6H_5O_7 \cdot 2H_2O$	0.1
	$MnSO_4 \cdot 7H_2O$	0.02
	Glucose	0.1
	Yeast extract	0.1

*LB: Luria-Bertani; GYS: glucose-yeast extract salt medium; SPY: Spizizen medium.

2.2 Preparation of H antisera and H agglutination studies

For H-serotype identification of the strain *mogi*, a slide agglutination test was used as described previously (Ohba and Aizawa, 1978). A motility inhibition test (Ishii and Ohba, 1993) was also involved in H-serotyping to confirm the specificity of the reaction. Reference antisera used were: (1) 55 H-antisera against the type strains of *B. thuringiensis* H1-H55 (Lecadet *et al*., 1999) (Table 2), and (2) four monospecific antisera against H-antigenic subfactors 3b, 3c, 3d, and 3e (Ohba and Aizawa, 1989) (Table 2-1). H antisera-antigen agglutination studies were performed using 96 well plates (Roh *et al*., 1996). One hundred microliter of flagellated bacteria suspension, grown at 30ºC to an OD_{600} of 0.7, was mixed in each well with 100 μl of H antiserum which had been diluted 50-fold with saline. Agglutinin was assayed after incubation at 37ºC for 1 h.

Table 2. H agglutination test results of *B. thuringiensis* subsp. *mogi* strain.

H-serotype	Serovar	*mogi*	H-serotype	Serovar	*mogi*
1	*thuringiensis*	-	25	*coreanensis*	-
2	*finitimus*	-	26	*silo*	-
3a3b3c	*kurstaki*	+	27	*mexicanensis*	-
3a3c	*alesti*	+	28a28b	*monterrey*	-
3a3d	*sumiyoshiensis*	+	28a28c	*jegathsan*	-
3a3d3e	*fukuokaensis*	+	29	*amagiensis*	-
4a4b	*sotto*	-	30	*medellin*	-
4a4c	*kenyae*	-	31	*toguchini*	-
5a5b	*galleriae*	-	32	*cameroun*	-
6	*entomocidus*	-	33	*leesis*	-
7	*aizawai*	-	34	*konkukian*	-
8a8b	*morrisoni*	-	35	*seoulensis*	-
8a8c	*ostriniae*	-	36	*malaysiensis*	-
8b8d	*nigeriensis*	-	37	*andaluciensis*	-
9	*tolworthi*	-	38	*oswaldocruzi*	-
10	*darmstadiensis*	-	39	*brasiliensis*	-
11a11b	*toumanoffi*	-	40	*huazhongensis*	-
11a11c	*kyushuensis*	-	41	*sooncheon*	-
12	*thompsoni*	-	42	*jinghongiensis*	-
13	*pakistani*	-	43	*guiyanggiensis*	-
14	*israelensis*	-	44	*higo*	-
15	*dakota*	-	45	*roskildiensis*	-
16	*indiana*	-	46	*chanpaisis*	-
17	*tohokuensis*	-	47	*wratislaviensis*	-
18	*kumamotoensis*	-	48	*balearica*	-
19	*tochigiensis*	-	49	*muju*	-
20a20b	*yunnanensis*	-	50	*navarrensis*	-
20a20c	*pondicheriensis*	-	51	*xiaguangiensis*	-
21	*colmeri*	-	52	*kim*	-
22	*shandongiensis*	-	53	*asturiensis*	-
23	*japonensis*	-	54	*poloniensis*	-
24	*neoleonensis*	-	55	*palmanyolensis*	-

-, no response; +, agglutination.

Table 2-1. Monospecific antisera agglutination test results of *B. thuringiensis* subsp. *mogi* strain.

monospecific antisera*	*mogi*	monospecific antisera	*mogi*
3b	+	3d	+
3c	-	3e	-

-, no response; +, agglutination.

* The result was co-worked with Ohba and Aizawa at Japan.

2.3 Plasmid preparation and PCR

Plasmid DNA was extracted using the alkaline lysis method (Reyes-Ramirez and Ibarra, 2008) including a step involving lysozyme treatment. Each strain was cultured in 50 ml SPY medium to an optical density at 600 nm of 0.8 to 1.0 at 30°C and 250 rpm shaking. Vegetative cells were pelleted at 7000 rpm for 10 min at 4°C. Each pellet was resuspended in 20 ml cold TES buffer (30 mM Tris base, 5 mM EDTA, 50 mM NaCl; pH 8.0 adjusted with 3 N HCl) and centrifuged under the same conditions. Cells were resuspended in 2 ml lysis buffer (TES buffer containing 20% sucrose, 2 mg/ml lysozyme, and 1µl/ml of RNase from a 10 mg/ml stock solution) and incubated at 37°C for 90 min or until more than 90% spheroplast formation was achieved and monitored under a microscope. The spheroplast suspension was supplemented with 3 ml of 8% sodium dodecyl sulfate in TES buffer and incubated at 68°C for 10 min. Then 1.5 ml of 3 M sodium acetate (pH 4.8) was added, and the suspension was incubated at -20°C for 30 min. The suspension was centrifuged at 10, 000 rpm for 20 min at 4°C. The supernatant was translucent; if it was not, another centrifugation was done, and ultimately, if still required, it was filtered. Two volumes of cold absolute ethanol were added to the supernatant and incubated overnight at -20°C. Plasmid enriched DNA was pelleted at 10,000 rpm for 20 min at 4°C. Each pellet was dissolved in 100 µl Tris-EDTA (pH 8.0) (10 mM Tris-HCl, 1 mM EDTA) and stored

at -20°C until further use.

For *cry*-gene typing in the strain *mogi*, PCR tests were done according to the method of Lee (Lee *et al*., 2001). Twenty major *cry*/*cyt* genes primers (*cry1Aa, cry1Ab, cry1Ac, cry1B, cry1C, cry1D, cry1E, cry1F, cry1I, cry2A, cry3A, cry3B, cry3C, cry4A, cry4B, cry7A, cry9A, cry10A, cry11A, and cyt1A*) were synthesized for PCR analysis. The purified PCR products were ligated to pGEM-T easy vector (Promega Co., USA) for sequencing and analyzed by dye termination method in ABI 377 automated sequencer (Applied Biosystems, USA).

2.4 Pulsed-field gel electrophoresis

Separation and examination of large DNA fragments was performed with pulsed-field gel electrophoresis as described by Dean and Bazylinski (1999) in a CHEF-DRII system (Bio-Rad Ltd, Richmond, CA). Agarose gels were prepared at a concentration of 1% and electrophoresis was performed in 0.5 × TBE buffer (45 mmol/l Tris–HCl, 45 mmol/1 boric acid, 2.5 mmol/l EDTA, pH 8.2) at 14°C for 16 h. Mid-Range II PFG Markers (New England Biolabs Ltd, Ipswich, MA) were used as molecular weight markers. The field strength and pulse conditions were 6 V/ cm, and switch times ramped from1 to 25 s.

2.5 Transmission electron microscopy

Crystal morphology of the isolate was examined by phase-contrast microscopy and

transmission electron microscopy. For TEM sample, the *B. thuringiensis* cells were harvested prior to autolysis and washed with sterile water once. Cells were primary fixed with 2 ml fixation buffer (2% paraformaldehyde and 2% glutaraldehyde in 0.05 M sodium cacodylate buffer, pH 7.2) for 2 h at 4ºC, followed by three washes with 0.05 M sodium cacodylate buffer. After post fixation in 1% osmium tetroxide for 2 hours and two brief washes with ddH_2O, the samples were stained with 0.5% uranyl acetate overnight and dehydrated in increasing concentrations of ethanol. Then the specimens were embedded in Spurr's resin at 70ºC for 24 h. Sections were cut with an ultramicrotome (MT-X, RMC, Tucson, AZ, USA) and were stained with 2% uranyl acetate and Reynold's lead citrate. The cells were observed under a transmission electron microscope (Libra 120, Carl Zeiss, Germany).

2.6 SDS-PAGE and MS analysis

Parasporal inclusions were purified by the method of Thomas and Ellar (1983). For SDS-PAGE samples, cell were cultured on NA (nutrient agar) medium plate at 28ºC and harvested after autolysis. SDS-PAGE was performed on a 12% separating gel with 5% stacking gel. The gel was stained with 0.1% Coomassie brilliant blue (Sigma Co., St Louis, MO, USA). The stained protein bands were identified by nano-LC-ESI-IT MS analysis performed by Korea Basic Science Institute. For the

similarity search, we performed BLAST searches with the query of matched sequences found in the contig sequences database of *Bacillus* using the SEQUEST program (version 3.3.1, Thermo Electron Corporation, USA).

2.7 Southern hybridization

Southern hybridization to total plasmid DNA of *B. thuringiensis mogi* strain was performed according to the manufacturer's instruction (Boehringer Mannheim, Germany). Total plasmid DNAs of *mogi* were separated on 0.8% agarose gels. The gels were treated for 15 min in 0.25 N HCl and transferred to Hybond N^+ filters (Amersham Pharmacia, Biotech, Sweden) in 0.2 N NaOH as transfer buffer. PCR amplified *cry* genes (Table 3) from *B. thuringiensis* servoar *mogi* strain were used as probes and labeled with digoxigenin using a DIG DNA labelling kit (Boehringer Manheim Co., Germany). Prehybridization, hybridization, washing and detection procedures were followed as described by the manufactures.

2.8 Insects and toxicity assays

The mosquito larvicidal activities were assayed on 4th instar larvae of *Culex pipiens molestus* and *Culex pipiens pallens* (Diptera: Culicidae) which were grown in a container (35 × 25 × 3 cm) at 25ºC. Freeze-dried *B. thuringiensis* spores-crystal

complex were suspended in double-distilled water. Suspensions were diluted to 6 or 7 different concentrations in cups in a final volume of 100 ml. Bioassays were replicated three times using 30 4th *Culex pipiens* instar of per concentration. After 48 h of exposure at 25°C, dead larvae were counted. Statistical analysis of data and 50% lethal concentrations (LC_{50}) were performed with probit analysis (Russell *et al.*, 1977).

Table 3. Nucleotide sequences of primers used for amplification of the specific probe in southern blot.

Primer [a]	**Sequence (5'-3')**	**Target gene**
Fw-27p	ATGAATCCTTATCAGGATAAGAATGAA	*cry27Aa*
Re-27p	ATTCTGATCGTACGTATTATATCCTT	
Fw-39p	CCGGCTGCACATGTAACC	*cry39orf2*
Re-39p	GGTTACATGTGCAGCCGG	
Fw-20p	ACATGTAGAACAACTTATTCAAC	*cry20-like*
Re-20p	GTTCTAATCCTGAATCCCCTG	

3. Results

3.1 Characteristic of *B. thuringiensis mogi*

According to H-serotyping test, flagellated vegetative cells of the isolate were agglutinated with the H3 reference antiserum only (Table 2). In a further test to identify the subfactors, *mogi* was agglutinated with 3b and 3d antisera but non-reactive for 3c and 3e antisera (Table 2-1). It is clear from the results that the H-antigen of the strain *mogi* comprises three subfactors: 3a, 3b, and 3d. The subfactor 3a is an antigen commonly contained in all of the *B. thuringiensis* strains that belong to the serotype H3 (De Barjac *et al.*, 1981; Ohba and Aizawa, 1989). According to the current H-serotyping scheme for *B. thuringiensis* (Lecadet *et al*., 1999), the serotype H3 is divided into four subserotypes: 3a3c (serovar *alesti*), 3a3b3c (serovar *kurstaki*), 3a3d (serovar *sumiyoshiensis*), and 3a3d3e (serovar *fukuokaensis*). Thus, our present results create a new subserotype, 3a3b3d, designated serovar *mogi*. The serovar name is derived from ‘‘mosquito” in Korean.

The plasmid DNA pattern of the isolate was compared with the profiles of *B. thuringiensis* H3 type strains (Fig. 1) as well as *B*. *thuringiensis* mosquitocidal type strains (Fig. 1-1). In contrast to the complicated plasmid profiles of H3 serotype and mosquitocidal type strains, *B. thuringiensis* subsp. *mogi*, which contains a very simple pattern without visible small plasmid observed in these figures. For further study, pulse field gel electrophoresis was carried out to confirm the result. Fig 2 clearly indicated that, there were at least two megaplasmid (larger more than 194 kb) bands but none small plasmid harbored in this strain.

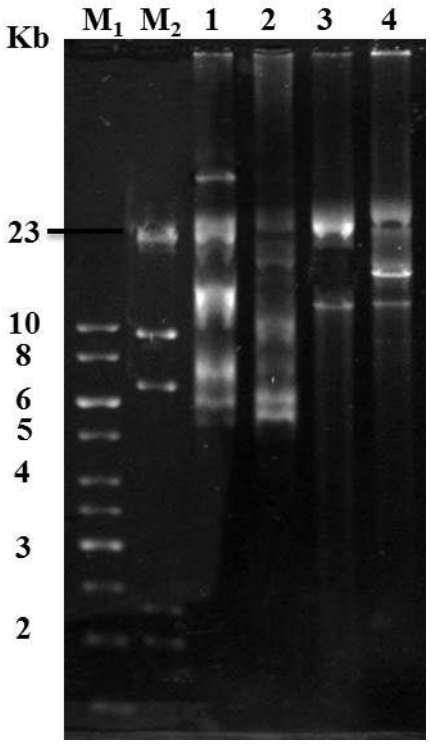


Fig. 1. Plasmid patterns from *B*. *thuringiensis* H3 type strains. The serovars are given, Lanes: 1, serovar *kurstaki* strain HD-1 (serotype 3a,3b,3c); 2, *alesti* (3a,3c); 3, serovar *sumiyoshiensis* (serotype 3a,3d); 4, *fukuokaensis* (3a,3d,3e). M_1, Gene Ruler^{TM} 1 kb DNA ladder; M_2, lambda DNA digested with *Hind* III.

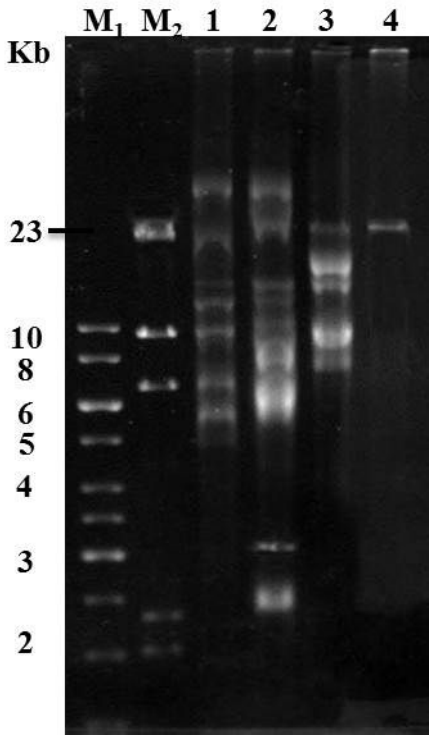


Fig. 1-1. Plasmid patterns from *B. thuringiensis* mosquitocidal type strains. The serovars are given, Lanes: 1, *israelensis* (H14); 2, *kyushuensis* (H11a11c); 3, *morrisoni* PG14 (H8a8b); 4, *mogi* (H3a3b3d). M_1, Gene Ruler^{TM} 1 kb DNA ladder; M_2, lambda DNA digested with *Hind* III.

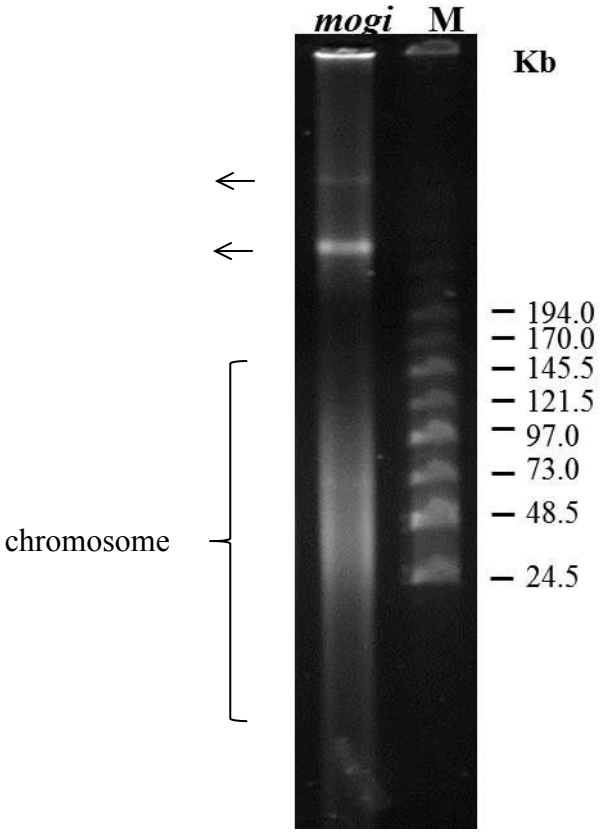


Fig. 2. Plasmid pattern from *B. thuringiensis* subsp. *mogi* strain was examined in pulse field gel electrophoresis. The Mid-Range II PFG Marker (M) was used as molecular weight markers (electrophoresis condition: 1% agarose gel, 6 V/cm, 15 for 18 h, switch times ramped from 1-25 s).

For detection of crystal genes of *B. thuringiensis* strains, PCR analysis was performed with *cry* gene-specific primers. The PCR-based identification of *B. thuringiensis cry* genes was first developed by Carozzi *et al*. (1991), who introduced this technique as a tool for prediction insecticidal activity. The PCR test with 20 specific primers failed to detect the genes allied to *cry1, cry2, cry3, cry4, cry7, cry9, cry10, cry11* and *cyt1*.

As shown in Fig. 3, the parasporal body is composed of 3 major endotoxins, a large ovoidal inclusion and two smaller inclusions combine together, with an distinct membrane outside the inclusions in one sporangium. The SDS–PAGE profile of the crystal proteins consisted of several bands ranging from 20 to 75 kDa (Fig. 4A). Fig 4B shows the internal amino acid sequences of putative Cry proteins obtained by nano-LC-ESI-IT MS analysis. Through the protein identification, three putative peptides of Cry39ORF2, Cry27Aa and Cry20-like were detected. Interestingly, the peptide No. 3 was evident in several fragments (Nos. 2, 3, 4, and 7) and the peptides Nos. 4, 5 and 7 were also found in two fragments. The existence of these putative peptides supports the observation that the mosquito-specific activity was associated with the strain *mogi*. Previous investigators reported the occurrence of mosquitocidal proteins Cry27Aa (Saitoh *et al*., 2000) in a serovar *higo* (H43) strain, and Cry20-like (Lee and Gill, 1997) in a serovar *fukuokaensis* (H3a3d3e) strain. Southern blot analysis showed that all of these three *cry* genes located in the total plasmid DNA of

strain *mogi* (Fig 5), this strain harbored three different *cry* genes at least.

3.2 Toxicity of *B. thuringiensis* subsp. *mogi*

The toxicity of wild-type *B. thuringiensis* subsp. *mogi* was evaluated against 4th instars of *Culex pipiens molestus* and *Culex pipeins pallens* larvae (Table 4). The spore-crystal mixture of the strain gave high mortalities of the two mosquito species, *C. pipiens molestus* (with an estimated LC_{50} of 16 μg/ml) and *C. pipens pallens* (with an estimated LC_{50} of 22.2 μg/ml). In contrast, it exhibited no larvicidal activity against three lepidopteran species: *Bombyx mori*, *Plutella xylostella* and *Lymantria dispar* (data not shown).

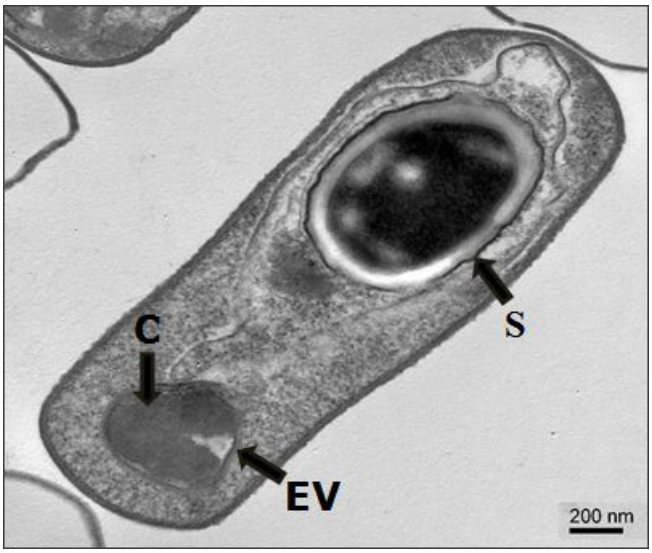
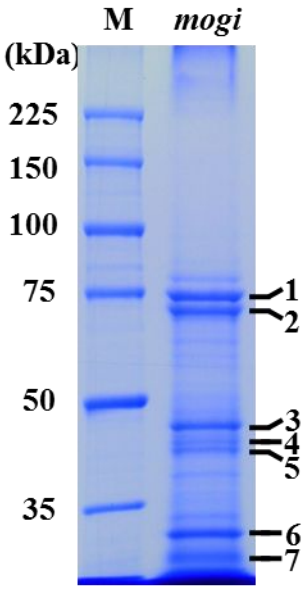


Fig. 3. Transmission electron microscopy of the parasporal crystal of *B. thuringiensis* subsp. *mogi* strain. Panel: C, S and EV indicate parasporal crystal, spore and envelope, respectively. Magnification is 60,000×.

A



B

Peptide No.	Internal sequence of tryptic peptide	Protein fragment No.	Best matched Protein	Predicted molecular mass	Serovar	GenBank Accession no.	Reference
1	K.YPLANDPQMY LR.N	1	Cry27Aa	94 kDa	*higo*	Q9S597	Saitoh et al. (2000)
2	K.TVEVFPESDRV R.I	2	Cry39ORF2	63 kDa	*aizawai*	BAB72017	-
3	R.IMQAYNLYDAR .N	2, 3, 4, 7					
4	K.AQLDGSGLAR .T	2, 3	Cry27Aa	94 kDa	*higo*	Q9S597	Saitoh et al. (2000)
5	R.YVPQISQVPAV K.A	2, 3					
6	K.ITTINLGDYDK. I	2					
7	R.SAATGAIYGIS R.S	5, 7	Cry20 like	86 kDa	*fukuokae nsis*	O32321	Lee and Gill (1997)

Fig. 4. SDS–PAGE and MS analysis of *B. thuringiensis* subsp. *mogi.*

A. SDS–PAGE profile of the parasporal crystal of *B. thuringiensis* subsp. *mogi* strain. Panel: M indicates the molecular marker.

B. Internal amino acid sequences of putative crystal proteins of the *B. thuringiensis* subsp. *mogi* identified by nano-LC-ESI-IT MS analysis.

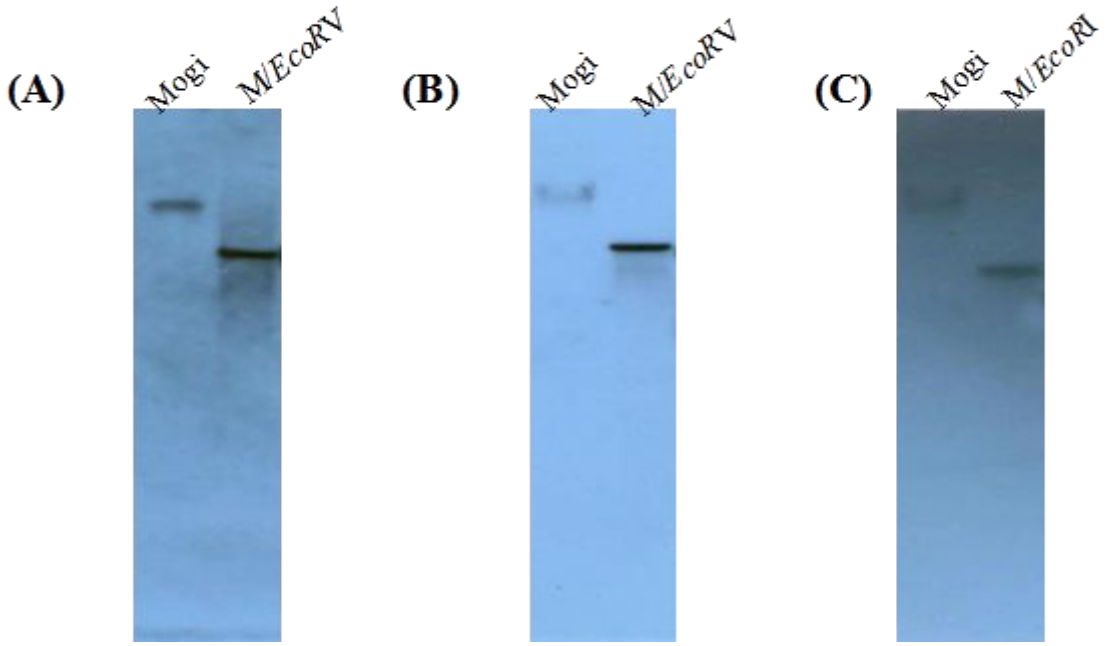


Fig. 5. Southern hybridization of *B. thuringiensis* subsp. *mogi* plasmid DNA with specific regions of *cry27Aa* (A), *cry39orf2* (B), and *cry20-like* (C) as probes , respectively. Lanes: Mogi, *B. thuringiensis* subsp. *mogi* plasmid DNA; M/*EcoR* V, *B. thuringiensis* subsp. *mogi* plasmid DNA digested with *EcoR* V; M/*EcoR* I, *B. thuringiensis* subsp. *mogi* plasmid DNA digested with *EcoR* I.

Table 4. Toxicity of *B. thuringiensis* against *Culex pipiens molestus* and *Culex pipiens pallens* 4th instar larvae.

Strain	***Culex pipiens molestus***		***Culex pipiens pallens***	
	LC_{50}[a] (µg/ml)	FL_{95}[b] (µg/ml)	LC_{50}[a] (µg/ml)	FL_{95}[b] (µg/ml)
mogi	16.0	14.1-21	22.2	13.7-27.5
israelensis	3.4	2.1-5.3	4.0	3.4-6.4

[a]LC_{50}: 50% lethal concentration (in µg) of freeze-dried spore–crystal complex per milliliter after 48 hours. The data are the total of three assays as determined by Probit analysis.

[b]FL_{95}: fiducial limits at P=0.95.

4. Discussion

B. thuringiensis is a gram-positive soil bacterium characterized by its ability to produce parasporal inclusions during sporulation. A number of *B. thuringiensis* isolates of the bacterium are commercially produced, with activity against Lepidoptera, Diptera and Coleoptera. Novel isolates with insecticidal activity have been recovered from numerous sources, particularly soil (Yamamoto and Powell, 1993), as well as grain dusts, diseased insect larvae, animal feed mills, phyloplane and aquatic environments (Coole, 1995). Meanwhile, *B. thuringiensis* shows great variability, as has been demonstrated by the huge number of strains isolated around the world (Xu *et al.*, 2013), by the number of serotypes known to date (a total of 84) (Roh *et al.*, 2009), and by the great number of different *cry* gene sequence accumulated so far (a total of 492), as well as by the number of molecular characterization tools that have been developed, such as sequencing of the *flagellin* gene and of the *gyrB* and *aroE* genes, the band patterns from repetitive extragenic palindromic-PCR analyses, and the plasmid patterns, among others (Choi *et al*., 2012; Heo *et al*., 2012; Koo *et al*., 2012; Reyes-Ramirez and Ibarra, 2008), all indicating the great variability within this species.

In this study, the characterization of a novel serogroup *B. thuringiensis* strain was reported. Characterization was based on serotype, plasmid pattern, crystal inclusion,

Cry protein composition, *cry* gene content and insect toxicity.

The diversity in flagellar H3 antigen agglutination reactions is one indication of the enormous genetic diversity among *B. thuringiensis* isolates. The plasmid pattern from *B. thuringiensis* subsp. *mogi*, showed a much simpler profile than other type strains. Strains of *B. thuringiensis* usually exhibit complex plasmid profiles, with molecules ranging from 2 to more than 200 kb (Hoflack *et al.*, 1997). Plasmids play a crucial role in bacterial evolution and adaptation by mediation the horizontal exchange of genetic material and providing advantageous functions to their carrier. The *B. thuringiensis* subsp. *israelensis,* which is highly toxic to larvae of several dipteran aquatic insects, has been reported to contain up to 10 plasmids (González and Carlton, 1984). Interest has been predominantly focused on large molecule plasmids where most of the crystal protein genes are encoded on. However, some small plasmids have been ascribed as no functions other than maintenance. They are referred to as cryptic plasmid. There are only megaplasmids but no small plasmid harbored in *mogi* strain made the isolate very different.

Sporulation in *B. thuringiensis* is associated with high protease activity which coincides with the onset of crystal formation (Andrews *et al*., 1985). Proteases endogenous to *B. thuringiensis* have been described from the cysteine, metallo, and serine families of enzyme, which can degrade Cry proteins and affect insect toxicity. Carroll *et al* (1989) reported that there was a reduction in the size of *B. thuringiensis*

subsp. *tenebrionis* inclusion crystal proteins occurred during sporulation, and proteolysis was prevented by the addition of proteases inhibitors. Mosquitocidal proteins were also degraded in the crystal (Dai and Gill, 1993).

In general, Cry proteins are active against Lepidopteran (Cry1 of 130–140 kDa), both Lepidopteran and Diptera (Cry II of 71 kDa), Coleopteran (Cry III of 66–77 kDa) and Diptera (Cry IV of 125–145 and 68 kDa) larvae (Guz *et al*., 2005; Salehi *et al*., 2008).

In this study, characteristically, isolate possessed δ-endotoxins with molecular weights between 20 and 75 kDa, among which ~70 kDa was distinctly present signifying their spectrum of activity against Diptera. A few crystalline inclusions were composed of the small components of polypeptides of 20 - 45 kDa. Both the MS analysis and southern blots proved the strain *mogi* contained three Cry proteins (Cry27Aa, Cry39ORF2 and Cry20-like), and the size of these protoxin is around 63 kDa to 94 kDa. It is well known that *B. thuringiensis* produces endogenous proteases and their production may vary considerably among strains (Rukmini *et al*., 2000). The present of these small polypeptides indicated the protoxins maybe cleaved during sporulation phase, giving rise to the smaller size. The new isolate also showed toxicity against *C. pipiens molestus* and *C. pipiens pallens* in bioassay test. In conclusions, the novel serovar type strain, subsp. *mogi* (H3a3b3d), will be a good resource for screening mosquitocidal Cry proteins.

Chapter 2. Genome Sequencing Strategy and Sequence Analysis of *Bacillus thuringiensis* subsp. *mogi*

ABSTRCT

Bacillus thuringiensis belongs to the *Bacillus cereus* sensu lato group as well as *B. anthracis* and *B. cereus*. Plasmids from this group of organisms have been implicated in pathogenicity as they carry genes responsible for a variety of mammalian and entomo-pathogenic diseases. In this study, genome sequence of the novel serogroup of *B. thuringiensis* subsp. *mogi* (H3a3b3d) was determined. The 6.0 Mb genome of *B. thuringiensis mogi* contains three replicons as follows: a circular chromosome (5.40 Mb) encoding 5,652 predicted open reading frames (ORFs) and two mega-plasmids, pMOGI364 (364,564 bp) and pMOGI222 (222,348 bp). The G+C contents of these replicons ranged from 31.3% to 34.2% for pMOGI364 and pMOGI222, respectively, and did not significantly deviated from that of the chromosome (35.2%). About 200 kb sequence of pMOGI364, showed a high similarity (more than 90% identity) to the plasmid pG9842_209 of *B. cereu*s G9842, and the last 146 kb fragment of pMOGI364 was found to harbor nine *cry* genes. The analysis of the replication-related sequence suggests that pMOGI222 may belong to the pAMβ1 family of Gram-positive theta-replicating plasmids. These sequences possibly contribute to the expansion of the pathogenic *B. thuringiensis* plasmid gene pool.

Key words: *Bacillus thuringiensis*, genome sequencing, megaplamid, *cry* gene, theta replicating mode

1. INTRODUCTION

Members of the *Bacillus cereus* group of organisms include *B. cereus, B. anthracis* and *B. thuringiensis*. This group of Gram-positive spore-formers forms a highly homogeneous subdivision of the genus *Bacillus*. The presence of Cry protein crystals in the spore is speculated to give *B. thuringiensis* an advantage in the soil environment upon sporulation (Jensen *et al*., 2003) over *B. cereus*, *B. thuringiensis* is phenotypically distinguished from *B. cereus* only by the formation of intracellular protein crystals during sporulation. Overall, genetic studies have shown that *B. cereus* and *B. thuringiensis* are essentially identical (Helgason *et al*., 1998).

The presence of a complex arrangement of plasmid DNA is a common characteristic of many strains in *B. thuringiensis*. The number and size of these plasmids (2–250kb) vary considerably among strains (González and Carlton, 1980). Often, plasmids confer an obvious advantage to the host, or encode traits that favor their own maintenance and survival. For many plasmids, however, no functions other than maintenance have been ascribed. They are referred to as cryptic plasmids.

In a plasmid pattern, two different groups of plasmids can be recognized: those that

are smaller than 30 MDa and those that are larger 30 MDa, called megaplasmids (Reyes-Ramirez and Ibarra, 2008). Small plasmids are generally present in high copy numbers, while megaplasmids are present in low copy numbers. As for the megaplasmids, their main recognized function is harboring *cry* genes, although the sequencing of some of these plasmids indicates the occurrence of other important genes (Berry *et al*., 2002; Chao *et al*., 2007; Jensen *et al*., 1995).

In addition, small plasmids generally use the rolling-circle replication mechanism, with single-stranded DNA intermediates, while megaplasmids normally use the "theta" replication mechanism (Wilcks *et al*., 1999). Theta replicons are currently divided into six groups (Group A-F) (http://www.essex.ac.uk/bs/staff/osborn/DPR/DPR_ThetaData.htm). Although there have been relatively few studies focusing on the characterization of Gram-positive theta replicons, as opposed to their Gram negative counterparts, plasmids pertaining to the broad host- range pAMβ1 family (group D) have been mostly studied from Gram-positive bacteria (Braund *et al.*, 1993; Brantl *et al.*, 1990; Swinfield *et al.*, 1990).

To date, five plasmids from the *B. cereus* group have been analyzed and reported to belong to the pAMβ1 family. Of these, the largest detected plasmid pBMB165 (about 82 kb) from *B. thuringiensis* subsp. *tenebrionis* YBT-1765, its mini-replicon has been determined (Huang *et al.,* 2006). p43 (65 kb) comes from *B. thuringiensis* subsp. *kurstaki* HD263, and a 2,828 bp replication region of p43 has been cloned (Baum and

Gilbert, 1991). The broad-host-range conjugative plasmid pAW63 (71,777 bp) has been isolated from *B. thuringiensis* subsp. *kurstaki* HD73, and a 4.1 kb replicon of pAW63 has been characterized (Wilcks *et al*., 1999). pBT9727 (77,112 bp) was the sole plasmid in the pathogenic strain *B. thuringiensis* subsp. *konkukian* 97-27, and its replication protein and the predicted origin have been analyzed by sequence comparison (Rasko *et al*., 2005). pXO2 (96,231 bp) was the second virulence plasmid of *B. anthracis*, and a 2,429 bp replication region has been identified (Tinsley *et al.*, 2004).

For *B. thuringiensis* genomics, there are 11 complete and 19 in-progress genomes publicly available on NCBI (http://www.ncbi.nlm.nih.gov/genome/486, as of November 12, 2013). *B. thuringiensis* strains have a genome size of 5.31 to 6.77 Mb. Here, the complete genome from an environmental isolate of *B. thuringiensis* subsp. *mogi* was determined by using shotgun libraries plus paired-end library sequencing strategy. The use of both libraries showed a more adequate representation of contigs and permitted the closure of the genome sequences. Also a comparative analysis with the genome of other *Bacillus* was proceeded and these data provide an insight into evolutionary relationships among the *Bacillus*.

2. MATERIALS AND METHODS

2.1 Bacterial strains and growth media

The novel serogroup *B. thuringiensis mogi* (H3a3b3d) strain used in this research, was isolated from fallen leaves, sampled in a forest region of the city of Mungyeong, as previously described (chapter 1) (Roh *et al.*, 2009). The LB medium was used as a primary culture of *B. thuringiensis* and the second culture of *B. thuringiensis* was grown at 28ºC with vigorous shaking in SPY medium for DNA preparation.

2.2 Plasmid DNA extraction and sequencing strategy

Plasmid DNA of *B. thuringiensis* subsp. *mogi* was isolated according to the manufacturer's protocols of QIAGEN midi prep. kit (QIAGEN Co., Germany) with an additional lysozyme treatment. The total 70 μg plasmid DNA of *B. thuringiensis mogi* was used to construct three libraries: (i) a GS FLX + shotgun library using the GS FLX + library preparation kit, (ii) an 8 kb-long paired-end library using the GS FLX paired-end kit, (iii) a HiSeq DNA shotgun library using the HiSeq2000 shot gun library kit. The libraries were sequenced using the Roche/454 pyrosequencing method on a Genome Sequencer FLX system (Macrogen, Korea) or the HiSeq™ 2000 platform (Illumina,San Diego, USA). In total, 64,395,859 and 40,850,371 bases were analyzed in single and paired-end reads, which yielded 10 million nucleotides

covering the genome ~16-fold (Table 5). Meanwhile, 1034 contigs were produced in 201 scaffolds through GS *de novo* assembler v2.6 (454 sequencing system software; Roche). Gaps within and between the scaffolds were confirmed and closed using primer walks and long-distance PCR amplification. End-sequencing of amplicons was carried out on an ABI 3730xl DNA Analyzer (Life Technologies). The complete sequences of chromosome (5,420,908 bp), pMOGI364 (364,564 bp) and pMOGI222 (222,348 bp) were determined.

2.3 Sequence annotation and analysis

Coding genes and pseudogenes across the genome were predicted using Glimmer (Delcher *et al.*, 1999), GeneMarkHMM (Lukashin and Borodovsky, 1998), and Prodigal (Hyatt *et al.*, 2010) and annotated by comparison with the NCBI-NR (Benson *et al.*, 2008). tRNA and rRNA were identified using tRNAscan-SE and RNAmmer, respectively. The annotation results were verified using Artemis (Rutherford *et al.*, 2000) and corrected manually gene by gene. Sequence similarities were determined using standalone BLAST programs (Altschul *et al.*, 1997) to search nucleotide and non-redundant protein databases from GenBank. Circular diagrams of plasmids were created using CGView server (http://stothard.afns.ualberta.ca/cgview_server/). Comparisons among related plasmids were made using BLAST programs

and multiple sequence alignments were performed using ClustalX (Thompson *et al.*, 1997).

Important *B. thuringiensis*-related sequences were also collected, including *cry*, δ endotoxin genes and mosquitocidal toxin for plasmid annotation and for surveying insecticidal genes.

3. RESULTS

3.1 General features of the genome sequence

The genome of *B. thuringiensis mogi* consists three replicons: a circular chromosome (5,420,908 bp) encoding 5,652 predicted open reading frames (ORFs), and two megaplasmids, pMOGI364 (364,564 bp), pMOGI222 (222,348 bp) (Table 6 and 7). The G+C content of the chromosome is 35.3%, while that of the plasmids are 31.3% and 34.2% , respectively. A total of 5,511 CDSs were identified in the chromosome. There are 102 tRNA genes representing all the 20 amino acids and 13 rRNA operons in the chromosome.

Comparison with *B. cereus* chromosomal maps suggests that all of these chromosomes have a similar organization in the half near the replication origin while displaying greater variability in the terminal half (Carlson *et al.*, 1996). The likely origin of replication of the chromosome of *B. thuringiensis mogi* was identified by

similarities to several features of the corresponding regions in *B. cereus* and other bacteria, including *dnaA* (chromosomal replication initiation protein, CDS g_0001mp) and *recF* (CDS g_0004mp) near the origin, GC nucleotide skew [(G-C)/(G+C)] analysis (Fig. 6), and the presence of multiple *dnaA* boxes and AT-rich sequences immediately upstream of the *dnaA* gene. The deduced replication termination site of the chromosome is believed to be localized near 2.6 megabases (Mb), according to GC skew analysis (Fig. 6), and the coding bias for the two strands of the genome is for the majority of CDSs to be on the outer strand from 0 to ~2.6 Mb and on the inner strand from ~2.6 Mb to the origin.

To reveal similarities and differences among the *B. cereus* group, genome of *B. thuringiensis mogi* was assembled in a pair-wise fashion. By running the MUMmer module according to the instructions of MUMmer 3.2, the results showed that *mogi* with two of the reference genomes, *B. anthracis* str. *Ames* (NC_003997.3) and *B. cereus* G9842 (NC_011772.1) (Fig. 7A and 7B) exhibit a much better synteny. While the other example, a comparison between *B. thuringiensis mogi* and *kurstaki* str. HD73 (NC_020238.1) seems to be difference, an inversion located on the blue segment of the region 2.7 Mb - 3.4 Mb.

Table 5. Read status of *B. thuringiensis* subsp. *mogi* genome assemblies.

No. of reads	No. of bases	Assembled	Partial	Singleton	Repeat	Outlier	Too short
309,746	100,868,054	296,142	6,229	4,936	2,003	436	0

- No. of reads: the read used in the assembly computation.
- No. of bases: the read's bases used in the assembly computation.
- Assembled: the read is fully incorporated into the assembly.
- Partial: only part of the read was included in the assembly.
- Singleton: the read did not overlap with any other reads in the input.
- Repeat: the read deemed to be from repeat regions.
- Outlier: the read was identified by the GS De Novo Assembler as problematic.
- Too short: the read was too short to be used in the computation.

Table 6. General characteristics of chromosome and gene prediction.

location	Size (bp)	G+C (%)	CDS	tRNA	5S rRNA	16S rRNA	23S rRNA
chromosome	5,420,908	35.3	5511	102	13	13	13

Table 7. General characteristics of two plasmids from strain *B. thuringiensis* subsp. *mogi.*

Plasmid	Size (bp)	G+C (%)	CDS	Pseudogene	Total gene	Coding density[a]
pMOGI364	364,564	31.3	357	53	410	1.125
pMOGI222	222,348	34.2	215	27	242	1.088

[a] The coding density is expressed in gene/kb.

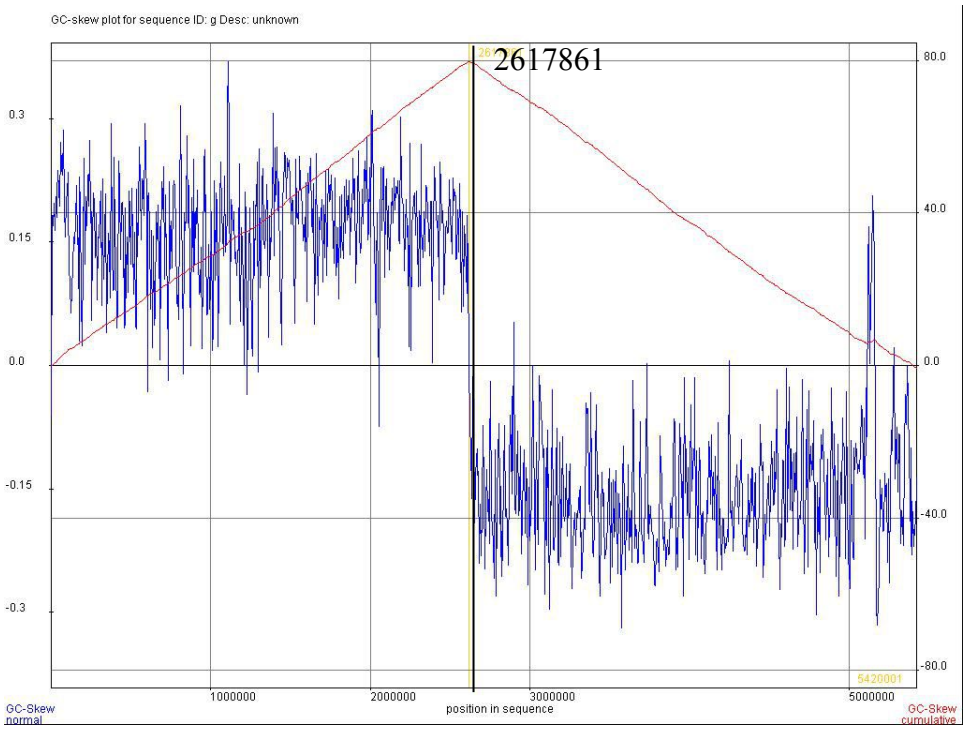


Fig. 6. GC-skew analysis of chromosome in *B. thuringiensis* subsp. *mogi*.

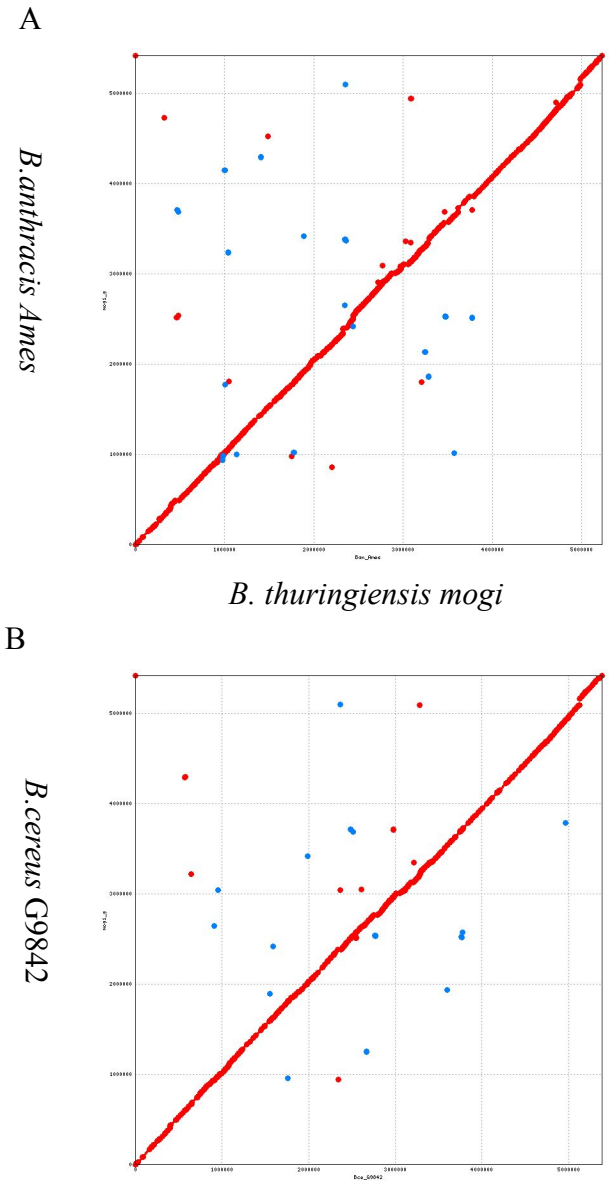
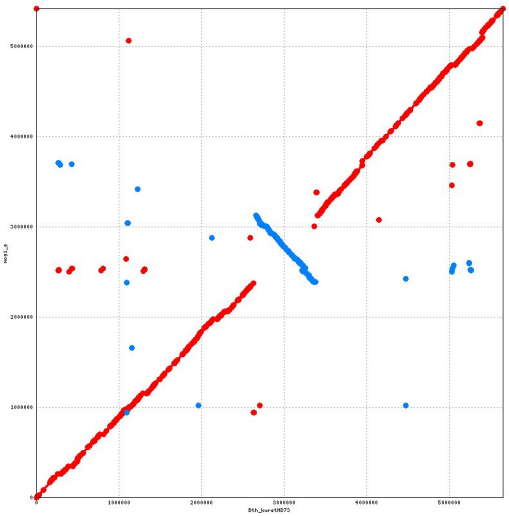


Fig. 7. Mummer analysis of *B. thuringiensis mogi* strain with the selected *Bacillus* species. (A) alignment of *B. thuringiensis mogi* to *B. anthracis Ames*; (B) alignment of *B. thuringiensis mogi* to *B. cereus* G9842. (C) alignment of *B. thuringiensis mogi* to *B. thuringiensis kurstaki* HD73. Matches in the forward strand are in red and those in the reverse strand are in blue. Linear regions are more conserved and scattered regions are less conserved.

C

B.thuringiensis kurstaki HD73



B. thuringiensis mogi

Fig. 7. Continued.

3.2 Phylogenic relationship between the *B. thuringiensis* subsp. *mogi* and the *B. cereus* group strains

The phylogenic tree was built based on 16S rRNA and DNA polymerase III α subunit (Zhao *et al*., 2007) sequences, using MEGA 5.2. Due to horizontal gene transfer, a major mechanism for genetic material exchange among prokaryotes, and the very close relations between members of *B. cereus* clade, which realized that phylogenic analysis may not yield stable relationship among the *B. cereus* group strains other than accurate and adequate selections of target sequences. Therefore, three sequences for this exercise: 16S rRNA, PolC, and DnaE, all of which are excellent for phylogeny were used. The latter two are DNA polymerase III α subunits of *Firmicute*s, whose sequences are highly conserved (Zhao *et al*., 2006; 2007). Although low bootstrap values indicated poor separations among the individual strains, the *B. thuringiensis mogi* strains showed much close to *B. cereus* in the 16S rRNA-based phylogeny (Fig. 8A). Furthermore, phylogeny based on PolC and DnaE (Fig. 8B) gave rise to a rather stable but fine differential relationship among all *B. cereus* group strains.

A

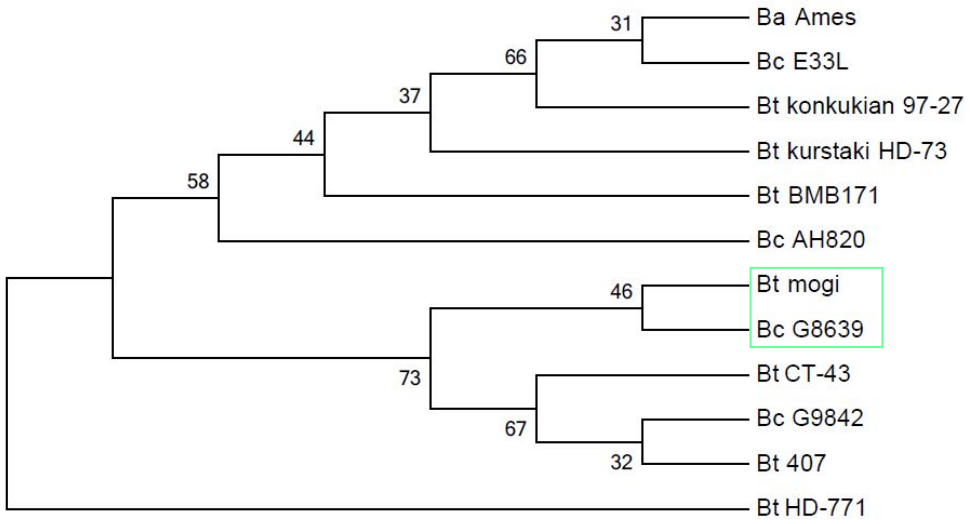


Fig. 8. Phylogenetic relationship of the *B. thuringiensis* subsp. *mogi* among the *Bacillus* clade. UPGMA trees were drawn with MEGA 5 following a sequential handling of selected CLUSTAL alignments coupled with bootstrapping. (A) Phylogenic survey based on 16S rRNA. (B) Phylogenic analysis of the isolates of *BacIllus* based on their PolC and DnaE genes.

B

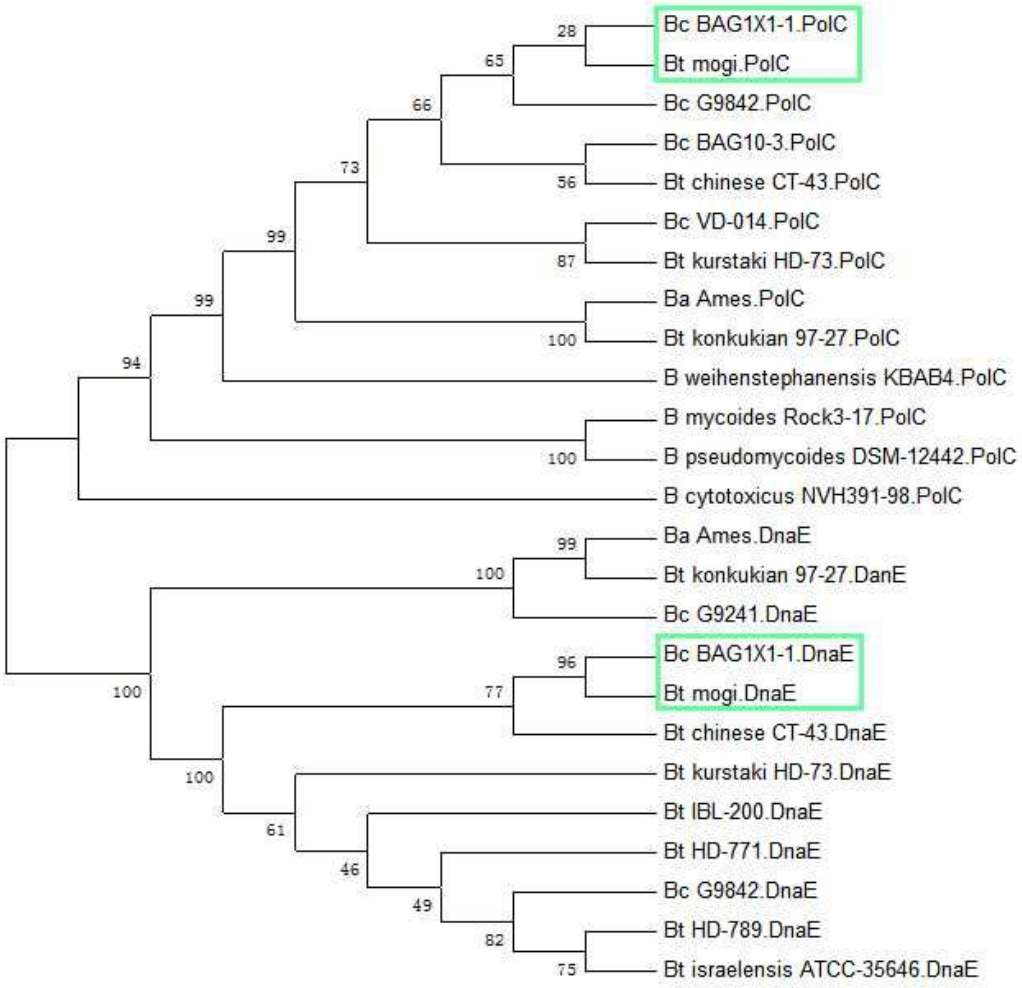


Fig. 8. Continued.

3.3 General features of plasmid sequence

Comparisons between the plasmid profiles of the novel serogroup *B. thuringiensis mogi* (H3a3b3d) and the H3 serotype (serovar *kurstaki*, *alesti*, *sumiyoshiensis* and *fukuokaensis*) showed that the novel serogroup isolate has a much simpler array of plasmids (Chapter 1, Fig. 1). This particular profile, and its lack of a relationship with type strains of the species, made the sequencing of extrachromosomal DNA present in *B. thuringiensis* subsp. *mogi* particularly interesting.

The complete nucleotide sequence of two plasmids from *B. thuringiensis* subsp. *mogi* was determined using two shotgun library plus 8 kb-long paired-end library sequencing method (see materials and method 2.2). The results and main characteristics of the plasmids are listed in Table 7. There are two megaplasmids, one approximately 364 kb, and one 222 kb plasmid. The plasmids were named pMOGI364 and pMOGI222 based on their sizes of 364,564 bp and 222,348 bp, respectively. All of them have G + C contents of around 30%, a value that is similar to that of plasmids from related species, such as *B. anthracis* and *B. cereus* (Andrup *et al*., 2003; Rasko *et al.*, 2005).

3.4 Plasmid pMOGI364

pMOGI364 properties are summarized in Table 7, and predicted genes are

described in Supplementary Table 1. A graphical representation of the plasmid is shown in Fig 11. pMOGI364 is the biggest plasmid from *B. thuringiensis* subsp. *mogi*. The complete DNA sequence was determined to be 364, 564 bp with a G+C content of 31.3%, within the range characteristic for the *B. thuringiensis* species. Of the 410 putative genes, 172 (42%) could be assigned putative functions, 186 (45.4%) encoded conserved hypothetical proteins, and 50 (12.2%) were pseudogenes.

(i) Analysis of the pMOGI364 sequence for the identification of the putative replicon

Most plasmids with sequenced genomes contain genes encoding a replication initiator protein. An alignment of such genes has shown that various plasmids can be grouped into plasmid families which share significant homologies in their replication initiator genes and the origin of replication (Chattoraj, 2000; del Solar *et al*., 1998; Khan, 1997). The genes encoding replication initiator proteins of both chromosome and plasmids are generally located in the vicinity of their replication origins. However, in the homology searches using a variety of software was failed to identify homologs of known replication initiator proteins in the pMOGI364 sequence. Then GC skew (Fig. 12), strand-specific biases such as gene orientation, plasmid-specific oligomer skew analysis, and origin comparisons (provided by the Genome Atlas Database at http://www.cbs.dtu.dk/services/GenomeAtlas/) were used to predict the location of an origin of replication (*ori*). The result suggested the possible location of the putative

ori of pMOGI364 was near nt 1 and the termination site was near nt 130,000.

(ii) FtsZ/tublin-related protein in pMOGI364

CDS pMOGI364_403, which shows high homology (99% Max identity) with FtsZ/tubulin-related protein, could potentially function in plasmid partitioning. The prokaryotic FtsZ protein is a polymer-forming GTPase that shares structural and functional similarities with eukaryotic tubulins (Vaughan *et al*., 2004). FtsZ assembles into a ring structure on the inner surface of the cytoplasmic membrane at the site of cell division. The so-called Z ring is progressively reduced in diameter, a process which leads to invagination of the dividing septum. Meanwhile, the CDS also shares limited homology (26% identity) with that encodes by pXO1-45 (*repX*), which was required for the replication of the miniplasmid in *B. anthracis* (Tinsley and Khan, 2006). In addition, ORF pBt156, which encodes a peptide with weak amino acid similarity to the FtsZ/tubulin-like proteins of *Pyrococcus* (BAB17294) and pXO1-45, are essential for replication of pBtoxis (Tang *et al*., 2006).

(iii) DNA topoisomerase III in *mogi*

Interestingly, pMOGI364 show similarity to *B. cereus* G9842 pG9842_209. About 200 kb sequence from pMOGI364, show a very high similarity (more than 90% identity) with the plasmid pG9842_209 (Fig. 13 and 14), and the last 146 kb fragment is found to harbor several *cry* genes. The level of protein similarity, combined with the conservation of gene order, suggests that these plasmids might

have diverged recently. pMOGI364 and pG9842_209 share a common backbone (Fig. 13). The insertion sites on pMOGI364 are between CDS pMOGI364_207 and pMOGI364_360, two CDS which combine together code for a intact type I DNA topoisomerase.

DNA topoisomerases resolve entangled DNA intermediates by transiently cleaving one or two DNA strands and passing another intact strand(s) through the nick. Based on their catalytic mechanism, topoisomerases have been categorized into four subfamilies, type IA and IB, and type IIA and IIB. Type IA topoisomerases are highly conserved from bacteria to humans. In most cells, two type IA topoisomerases are present while among the *Bacillus, B. anthracis* (Read *et al*., 2003), *B. cereus* (Ivanova *et al*., 2003) and *B. thuringiensis* have three chromosomal copies of type IA topoisomerases. Two plasmids, pXO1 and pXO2 which can be mobilized in *B. anthracis,* appear to encode two additional type IA topoisomerases. Therefore, five type IA topoisomerases may be present in these bacteria.

Here, DNA topoisomerase III from *B. thuringiensis* subsp. *mogi* are summarized in Table 8, there are five topoisomerase III on chromosomal and three on plasmid pMOGI364. Among them, four topoisomerase (g_0397mp, g_0413mp, g_1847mp, g_1858mp) from chromosomal are truncated and only one topoisomerase (g_0404mp) is intact. The same as topoisomerase on pMOGI364, two of them (pMOGI364_207, pMOGI364_360) are truncated and one (pMOGI364_233) is intact (Fig. 13). The

megalign results showed that the two neighbored truncated topoisomerase join together could be an intact topoisomerase III (g_0397mp plus g_0413mp, Fig 14-1; g_1847mp plus g_1858mp, Fig. 14-2; pMOGI364_207 plus pMOGI364_360, Fig. 14-3).

Prokaryotic type IA topoisomerases may have more diverse functions than simply the maintenance of genomic stability. These enzymes may have roles in the horizontal gene transfer of promiscuous plasmids or conjugational transposons. Type IA topoisomerases may also function in the process of conjugational DNA transfer, transposon integration (Sutanto *et al*., 2002), plasmid maintenance and plasmid segregation. To deal with stringent environmental stressors such as UV, chemical and free radical damage, bacteria may not only need sophisticated DNA damage repair systems, but also efficient DNA recombination systems to create, adopt and spread endogenous or exogenous mutations. Therefore, different type IA topoisomerases may be required in a variety of DNA repair and recombination processes. In addition, the characteristic cellular development stages such as sporulation and germination of spore-forming Gram-positive bacteria may require DNA replication and recombination systems as well (Li *et al*., 2005). These results predicted that the intact topoisomerase III on pMOGI364 may be essential for this strain.

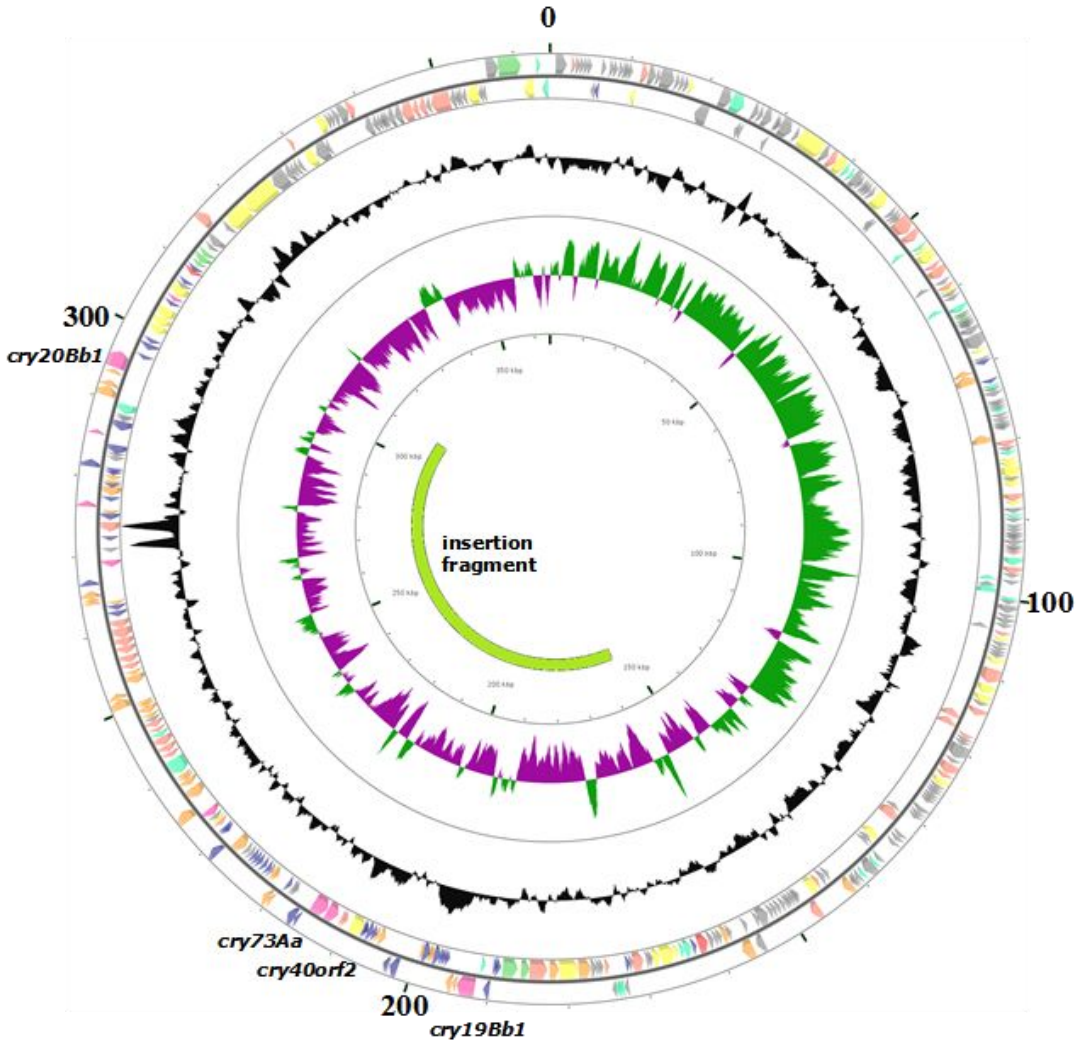


Fig. 9. Circular representation of pMOGI364. The inner circle represents GC bias

[(G- C)/(G+C)], with positive values in green and negative values in purple; the second circle represents GC content; and the outer two circles represent predicted genes on the reverse and forward strands (selected toxin-related CDSs are marked for reference). Predicted function/homologies are indicated by the color key featured below. The outer scale is marked in kilobases. The CDS number corresponds to Supplementary Table 1.

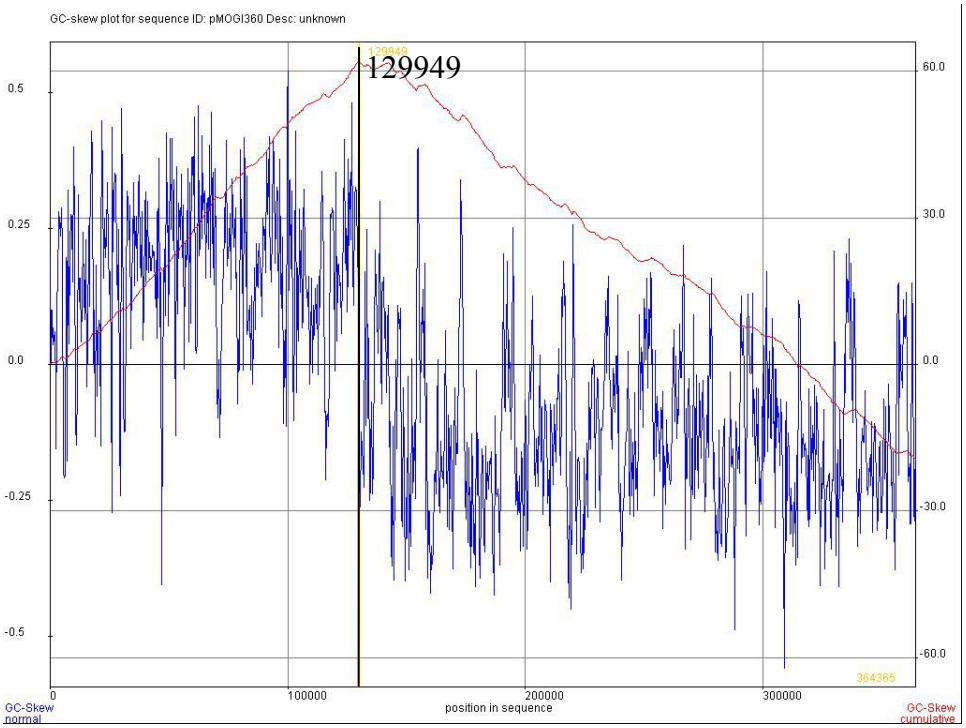


Fig. 10. GC-skew of the plasmid pMOGI364 in *B. thuringiensis* subsp. *mogi*.

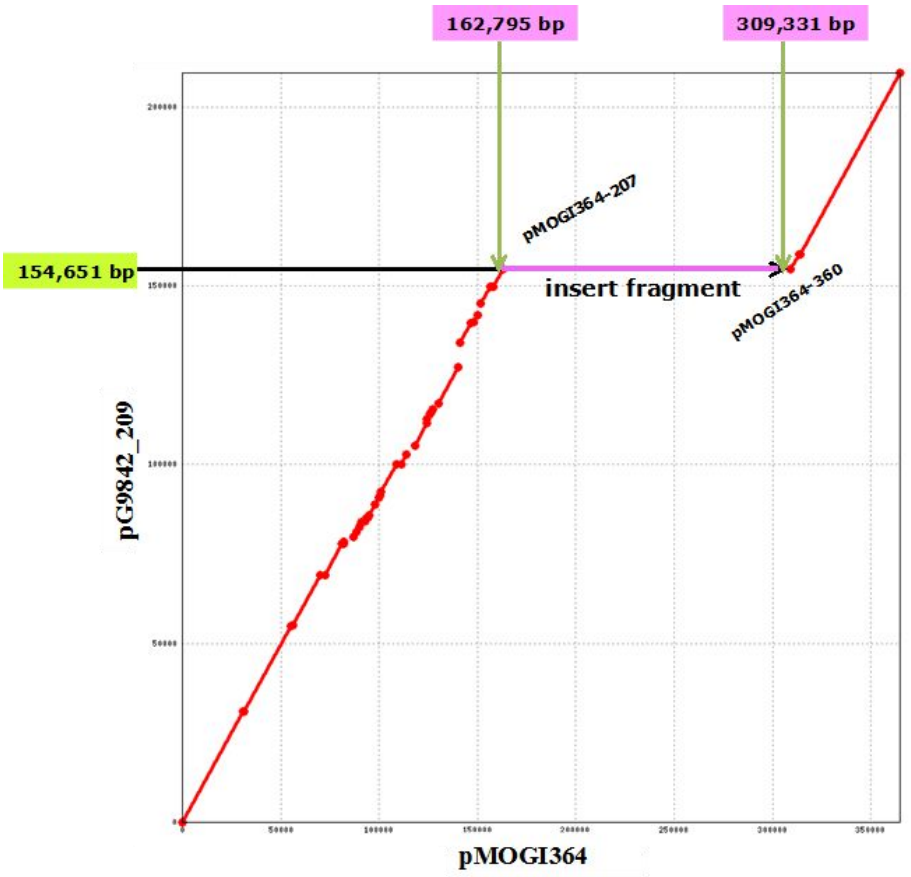
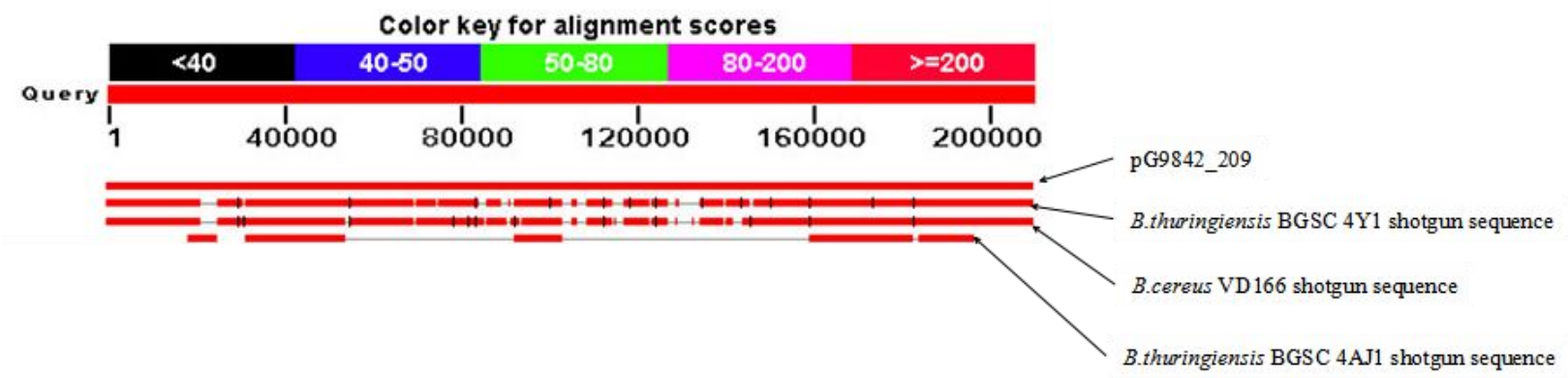


Fig. 11. Mummer analysis of pMOGI364 with pG9842_209.

(A)



(B)

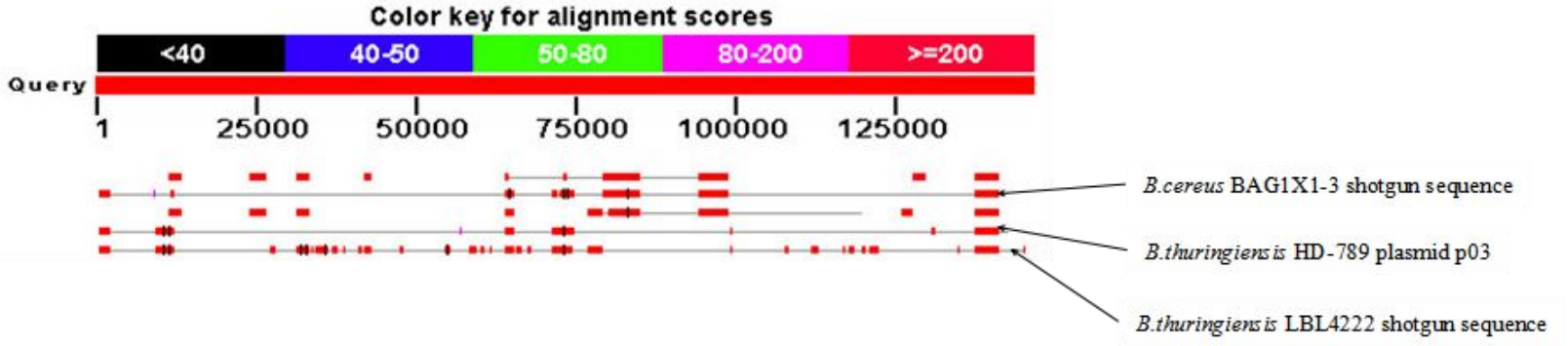


Fig. 12. Graphic overview of the results of megablast search using the sequence of pMOGI364 218 kb fragment (A) and 146kb insertion fragment (B) as a query against the database.

Table 8. DNA topoisomerase III in *B. thuringiensis* subsp. *mogi*.

No.	ltag	size	strand	predicted product homology with	(% aa identity)
1-1	g_0397mp	562 aa	-	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002443902.1) (729 aa)	98% in 543 aa
1-2	g_0413mp	198 aa	-	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002443902.1) (729 aa)	99% in 192 aa
2	g_0404mp	729 aa	+	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002443902.1) (729 aa)	90% in 730 aa
3-1	g_1847mp	369 aa	+	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002445288.1) (714 aa)	99% in 367 aa
3-2	g_1858mp	348 aa	+	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002445288.1) (714 aa)	99% in 348 aa
4-1	pMOGI360_207	446 aa	-	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002454832.1) (687 aa)	97% in 445 aa
4-2	pMOGI360_360	253 aa	-	DNA topoisomerase III [*Bacillus cereus* G9842] (YP_002454832.1) (687 aa)	96% in 243 aa
5	pMOGI360_233	716 aa	-	DNA topoisomerase III [*Bacillus cereus* HuB4-4] (EOP91371.1) (716 aa)	95% in 716 aa

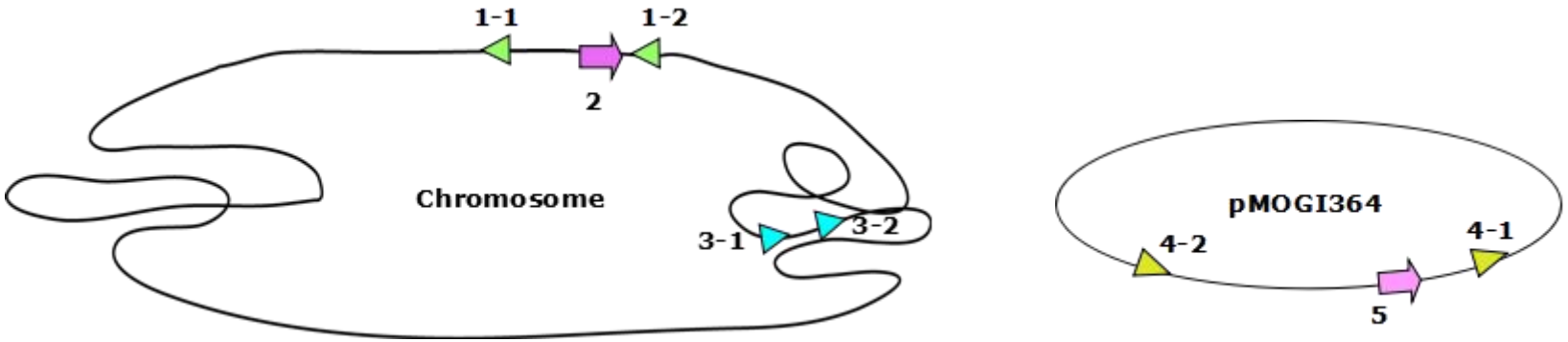


Fig. 13. Previews about DNA topoisomerase III in *B. thuringiensis* subsp. *mogi.* Truncated DNA topoisomerase III genes are showed in triangles and the intact DNA topoisomerase III genes are mapped in arrows.

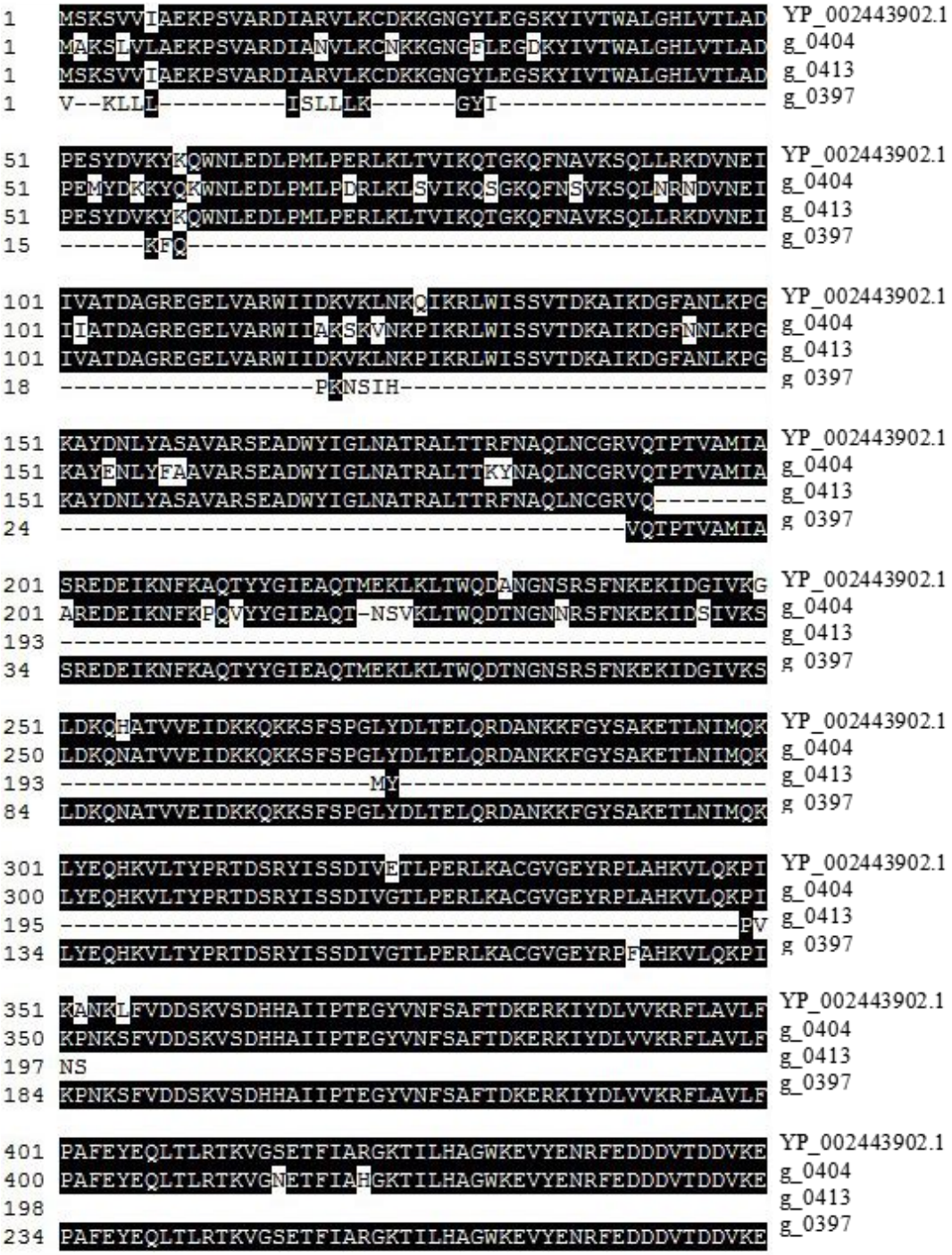
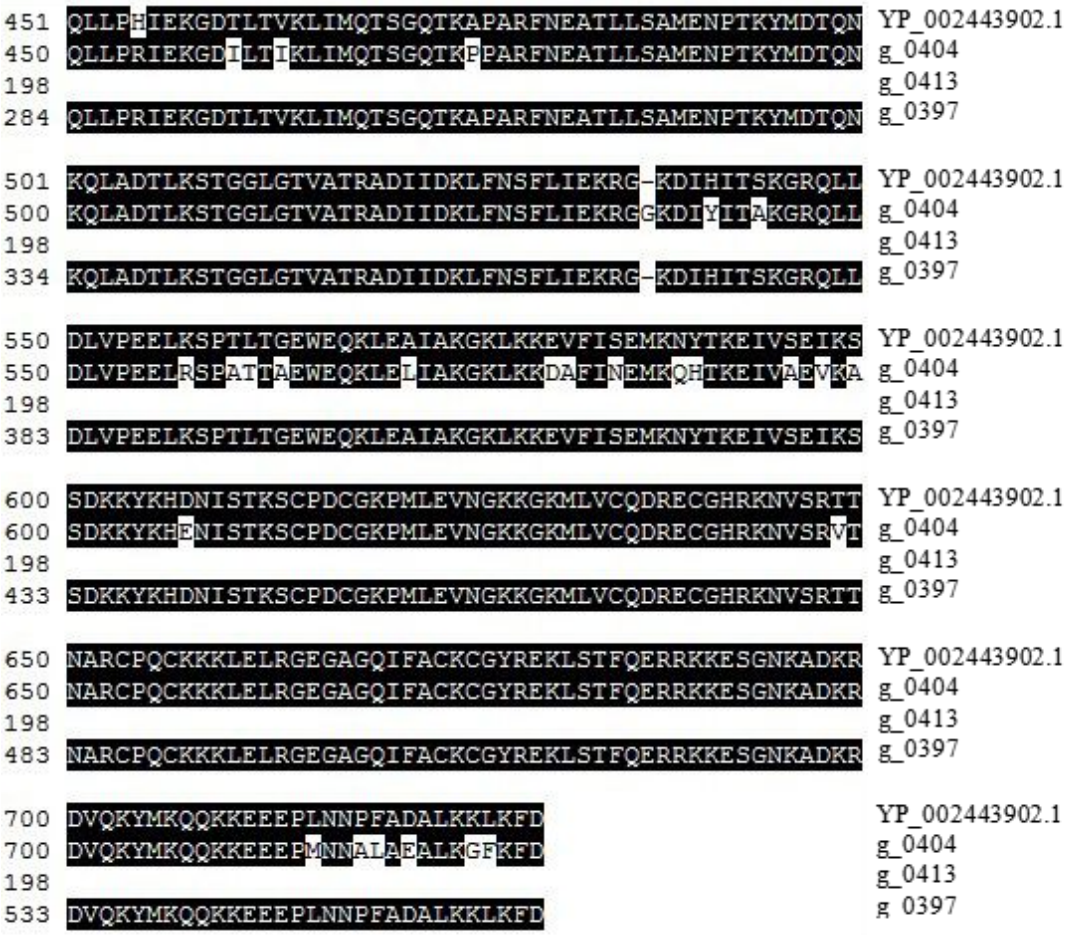


Fig. 14-1. Sequence alignment analysis of the truncated DNA topoisomerase with the intact one from NCBI database. Protein sequence comparisons were performed using the MegAlign program.

- continued


451 QLLPHIEKGDTLTVKLIMQTSGQTKAPARFNEATLLSAMENPTKYMDTQN YP_002443902.1
450 QLLPRIEKGDILTIKLIMQTSGQTKPPARFNEATLLSAMENPTKYMDTQN g_0404
198 g_0413
284 QLLPRIEKGDTLTVKLIMQTSGQTKAPARFNEATLLSAMENPTKYMDTQN g_0397
501 KQLADTLKSTGGLGTVATRADIIDKLFNSFLIEKRG-KDIHITSKGRQLL YP_002443902.1
500 KQLADTLKSTGGLGTVATRADIIDKLFNSFLIEKRGGKDIYITAKGRQLL g_0404
198 g_0413
334 KQLADTLKSTGGLGTVATRADIIDKLFNSFLIEKRG-KDIHITSKGRQLL g_0397
550 DLVPEELKSPTLTGEWEQKLEAIAKGKLKKEVFISEMKNYTKEIVSEIKS YP_002443902.1
550 DLVPEELRSPATTAEWEQKLELIAKGKLKKDAFINEMKQHTKEIVAEVKA g_0404
198 g_0413
383 DLVPEELKSPTLTGEWEQKLEAIAKGKLKKEVFISEMKNYTKEIVSEIKS g_0397
600 SDKKYKHDNISTKSCPDCGKPMLEVNGKKGKMLVCQDRECGHRKNVSRTT YP_002443902.1
600 SDKKYKHENISTKSCPDCGKPMLEVNGKKGKMLVCQDRECGHRKNVSRVT g_0404
198 g_0413
433 SDKKYKHDNISTKSCPDCGKPMLEVNGKKGKMLVCQDRECGHRKNVSRTT g_0397
650 NARCPQCKKKLELRGEGAGQIFACKCGYREKLSTFQERRKKESGNKADKR YP_002443902.1
650 NARCPQCKKKLELRGEGAGQIFACKCGYREKLSTFQERRKKESGNKADKR g_0404
198 g_0413
483 NARCPQCKKKLELRGEGAGQIFACKCGYREKLSTFQERRKKESGNKADKR g_0397
700 DVQKYMKQQKKEEEPLNNPFADALKKLKFD YP_002443902.1
700 DVQKYMKQQKKEEEPMNNALAEALKGFKFD g_0404
198 g_0413
533 DVQKYMKQQKKEEEPLNNPFADALKKLKFD g_0397

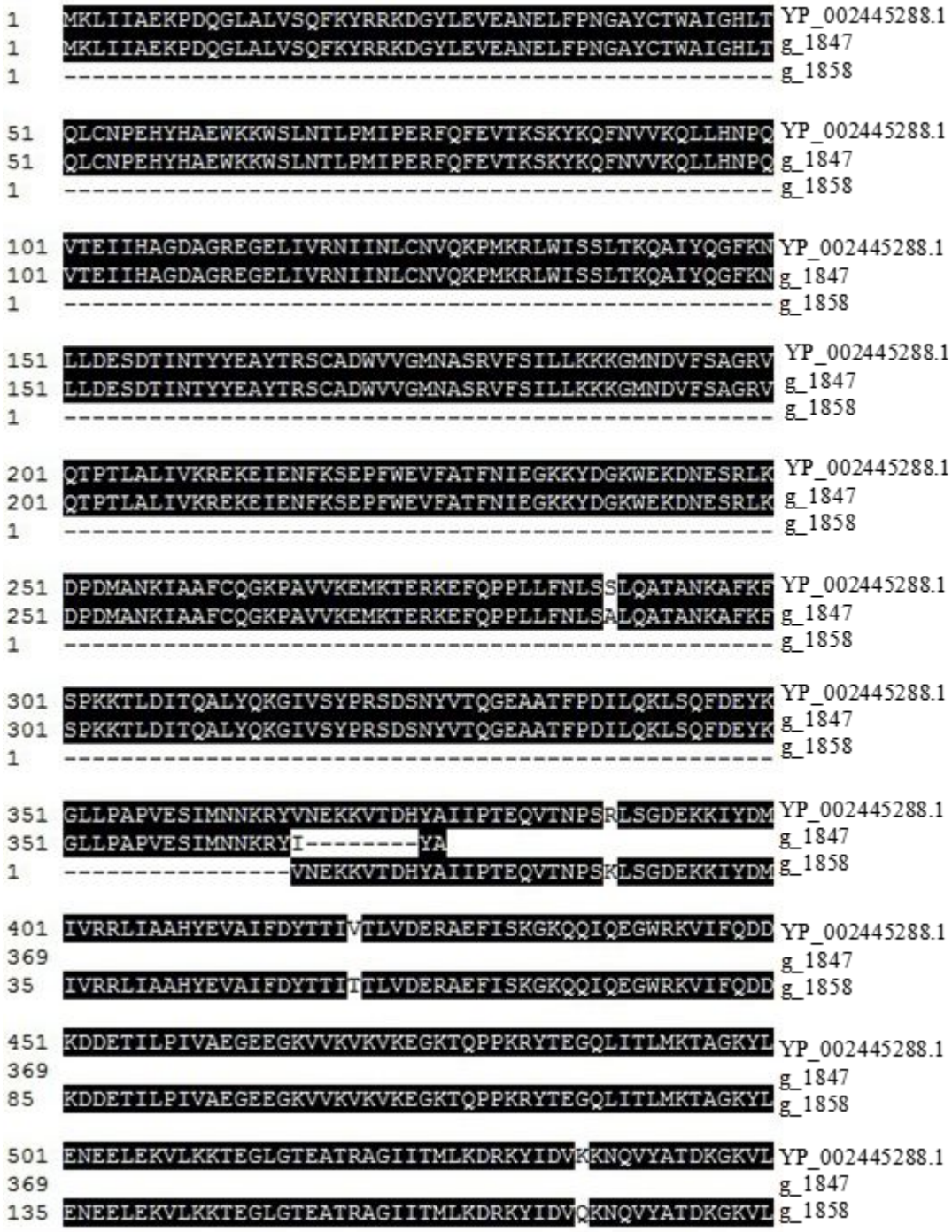


Fig. 14-2. Sequence alignment analysis of the truncated DNA topoisomerase with the intact one from NCBI database. Protein sequence comparisons were performed using the MegAlign program.

- Continued.


551 ITAIGDKILASPEMTAKWEQRLAEIGEGTASPATFMEQTKKLSAKIIEDA YP_002445288.1
369 g_1847
185 ITAIGDKILASPEMTAKWEQRLAEIGEGTASPATFMEQTKKLSAKIIEDA g_1858
601 VEMSEKWDFTGLHVESIERKGSKFTTGKKVGSCKKCDGDVIDKSTFYGCS YP_002445288.1
369 g_1847
235 VEMSEKWDFTGLHVESIERKGSKFTTGKKVGNCKKCDGDVIDKSTFYGCS g_1858
651 NYNTTQCDFTISKKILSKTISQKNMTKLLKGEKTDLIKGFKKGEKTFDAK YP_002445288.1
369 g_1847
285 NYNTTQCDFTISKKILSKTISQKNMTKLLKDEKTDLIKGFKKGEKTFDAK g_1858
701 LEWKDNKINFVFEN YP_002445288.1
369 g_1847
335 LEWKDNKINFVFEN g_1858

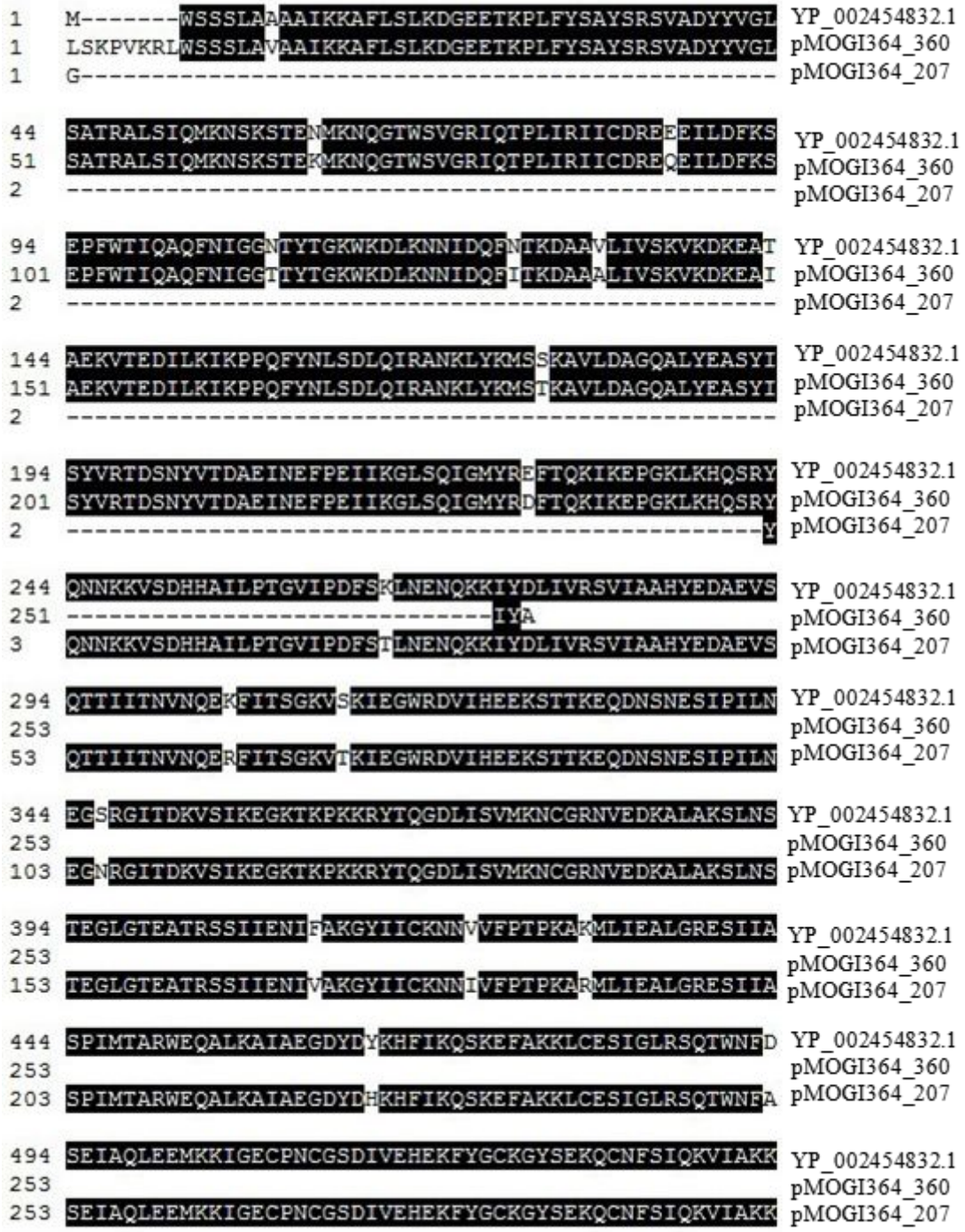


Fig. 14-3. Sequence alignment analysis of the truncated DNA topoisomerase with the intact one from NCBI database. Protein sequence comparisons were performed using the MegAlign program.

- continued


544 KISPAQVKKLLKDKKTDVIKGFKSSKSERTFETFLYYDSEKKCIDWGFNQ YP_002454832.1
253 pMOGI364_360
303 KISPAQVKKLLKDKKTDVIKGFKSSKSERTFETFLYYDSEKKCIDWGFNQ pMOGI364_207
594 AKNDNKPSSKDTGFKCPLCKDNLVEHQKFIGCSGFKNGCEFKISKNICGV YP_002454832.1
253 pMOGI364_360
353 AKNDNKPSSKDTGFKCPLCKNNLVEHQKFIGCSGYKNGCEFKISKNICGV pMOGI364_207
644 NLTPTHIEELVNNGETSMIDGFVFKDKTFRKALCVIDGKVMFKK YP_002454832.1
253 pMOGI364_360
403 NLTSTHIEELVNNGETSMIDGFVFKDNTFRKALCVIDGKVMFKK pMOGI364_207

3.5 Plasmid pMOGI222

The 222,348 bp sequence of plasmid pMOGI222 has 34.2% G + C content and 242 predicted genes. Twenty-seven (11.2%) of the predicted genes are pseudogene and 90 (37.2%) are hypothetical, whereas 7 CDS showed homology with toxin-related genes in NCBI database sequences, and 1 gene was thought to enhance crystal formation and subsequent cell viability by acting as chaperon (Supplementary Table 2). A graphical representation of the plasmid is shown in Fig. 15.

The protein proposed as responsible for the replication of pMOGI222 (Rep, replication initiation protein, pMOGI222_003) showed a high similarity to the Rep proteins of the pAMβ1 family of theta-replicating plasmids, such as Rep165 of pBMB165 (96.5% identity), Rep of pBT9727 (95.8% identity), Rep63A of pAW63 (82.5% identity) and RepS of pXO2 (81.1% identity) from *B. anthracis* (Tinsley *et al*., 2004), RepE of pAMβ1 (42.1% identity) from *Enterococcus faecalis* (Bruand *et al*., 1993) (Fig. 18). These conjugative plasmids replicate by a theta-type mechanism, and their replication proceeds unidirectionally from the origin (Bruand *et al*., 1993).

In addition, two cis-functioning regions were found near the Rep protein in pMOGI222. One cis-functioning region, located immediately downstream of Rep, displayed significant similarity to the cis-functioning origin of replication (*ori*) harbored in the corresponding locus of the pAMβ1 family plasmids cited above (Fig. 19). The other cis-functioning region was constituted of iterons (Fig. 20). The iteron

region was located between the copy number control gene (*parA*, pMOGI222_002) and the replication gene (*rep*, pMOGI222_003). These were sets of repeated DNA sequences that have been reported to serve as a binding site for the replication initiation protein and thus to play an important role in the replication and/or the control of replication in other theta-replicating plasmids. In addition, a AT-rich DNA region in the vicinity of the iteron sequences was found and marked with underline in Fig. 20.

Taken together, the conservation among these Rep proteins and *ori* regions thus provided significant evidence that pMOGI222 belongs to the pAMβ1 family of Gram-positive theta-replicating plasmids. All theta replication proteins from *B.cereus* group plasmids share a significant level of similarity and cluster together phylogenetically suggesting a common ancestral origin (Fig. 21).

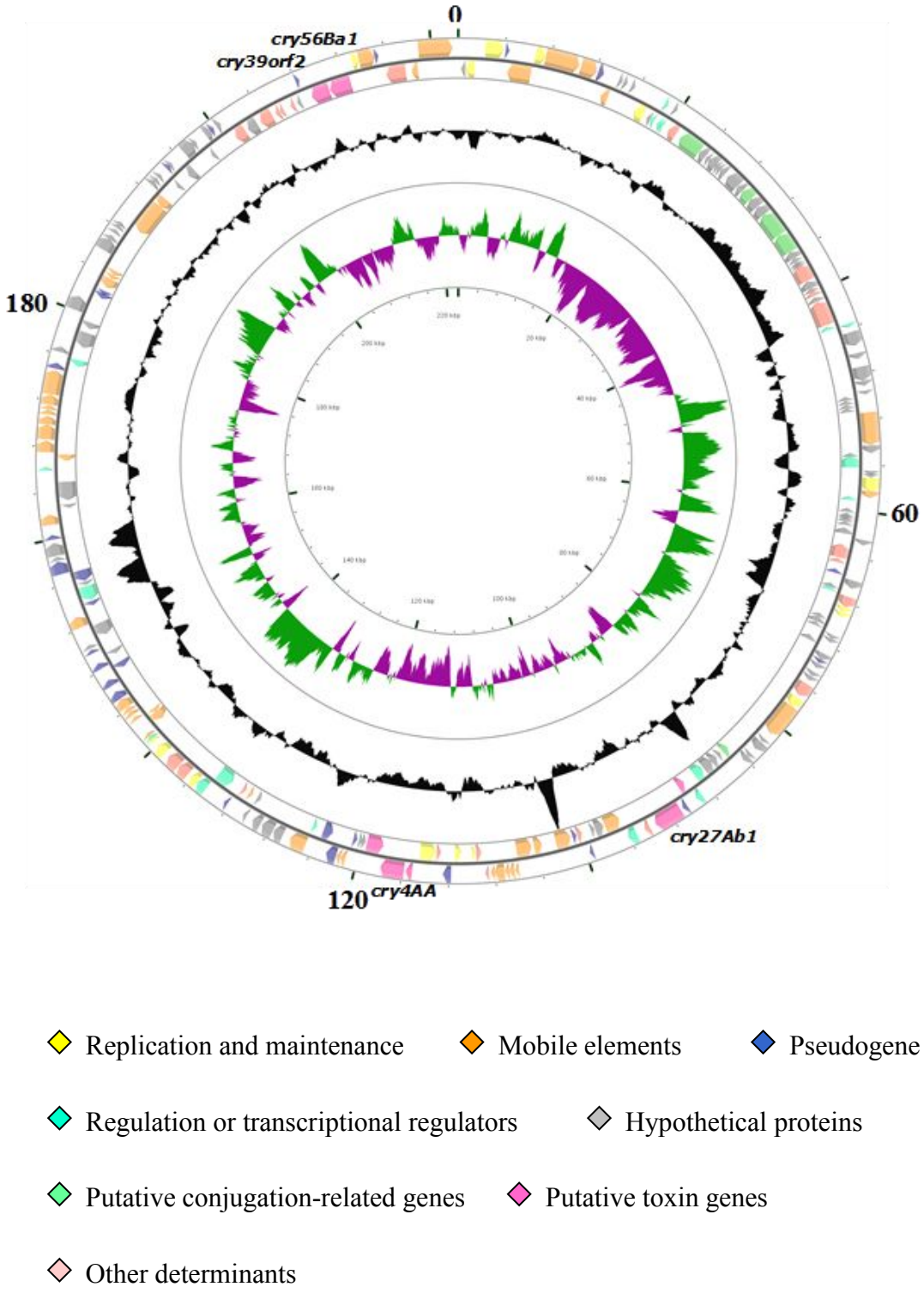


Fig. 15. Circular representation of pMOGI222. The inner circle represents GC bias

[(G- C)/(G+C)], with positive values in green and negative values in purple; the second circle represents GC content; and the outer two circles represent predicted genes on the reverse and forward strands (selected toxin-related CDSs are marked for reference). Predicted function/homologies are indicated by the color key featured below. The outer scale is marked in kilobases. The CDS number corresponds to Supplementary Table 2

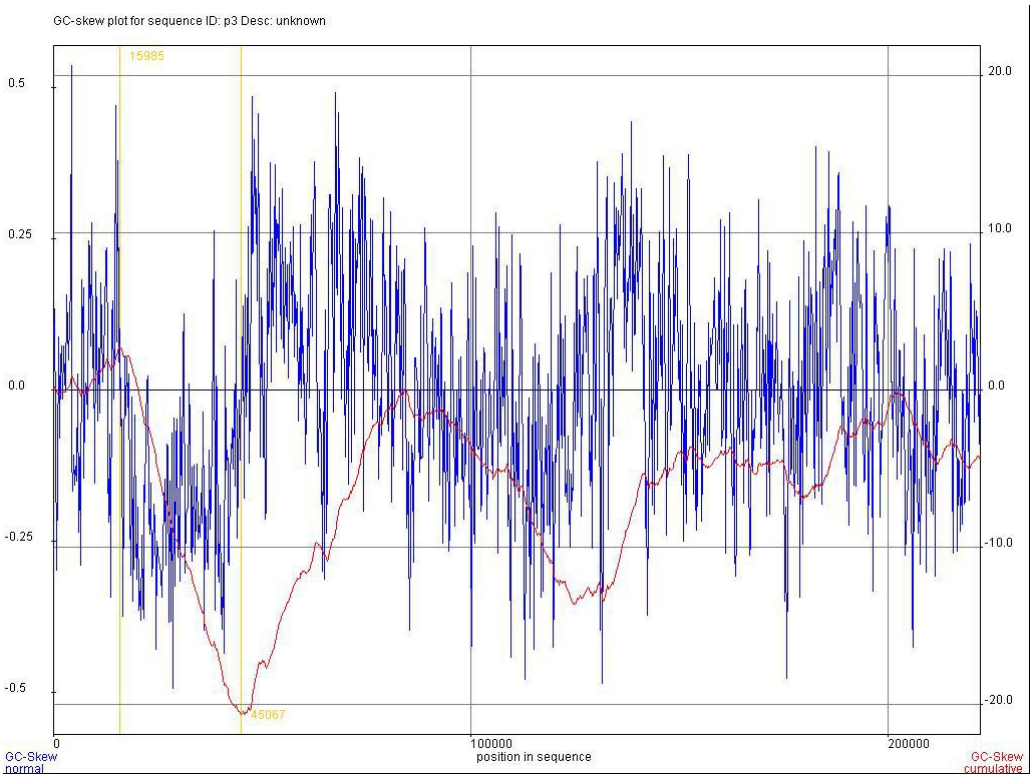


Fig. 16. GC-skew of the plasmid pMOGI222 in *B. thuringiensis* subsp.*mogi.*

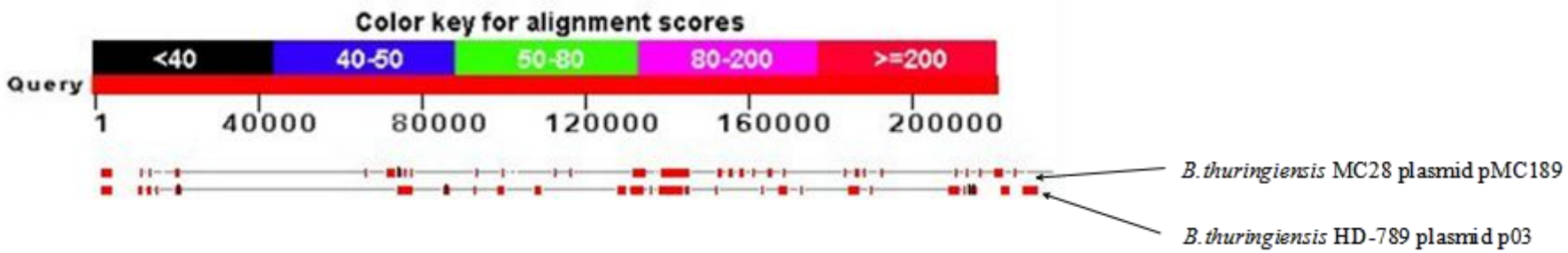
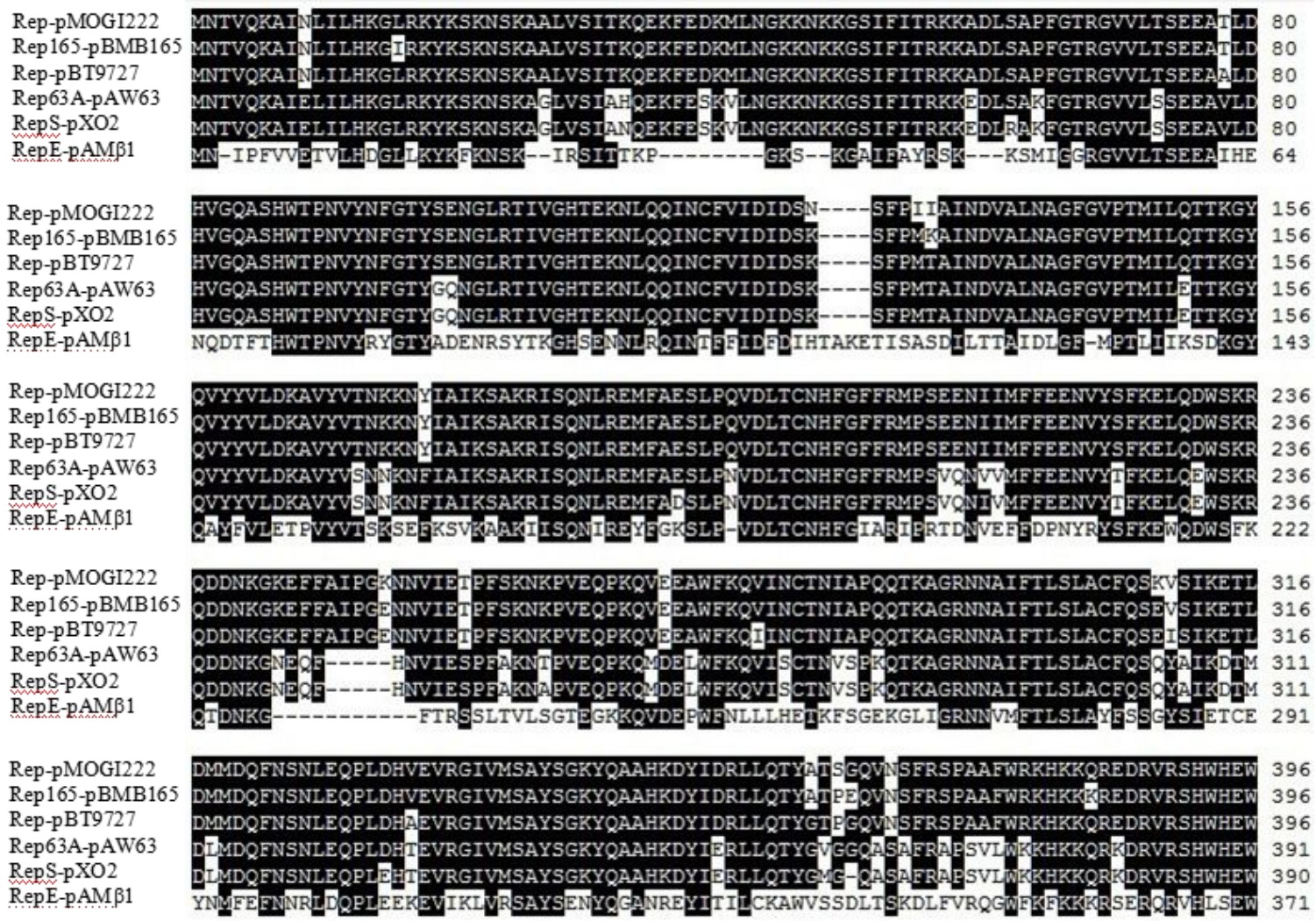


Fig. 17. Graphic overview of the result of a megablast search using the sequence of pMOGI222 as a query against the database.


Rep-pMOGI222
Rep165-pBMB165
Rep-pBT9727
Rep63A-pAW63
RepS-pXO2
RepE-pAMβ1

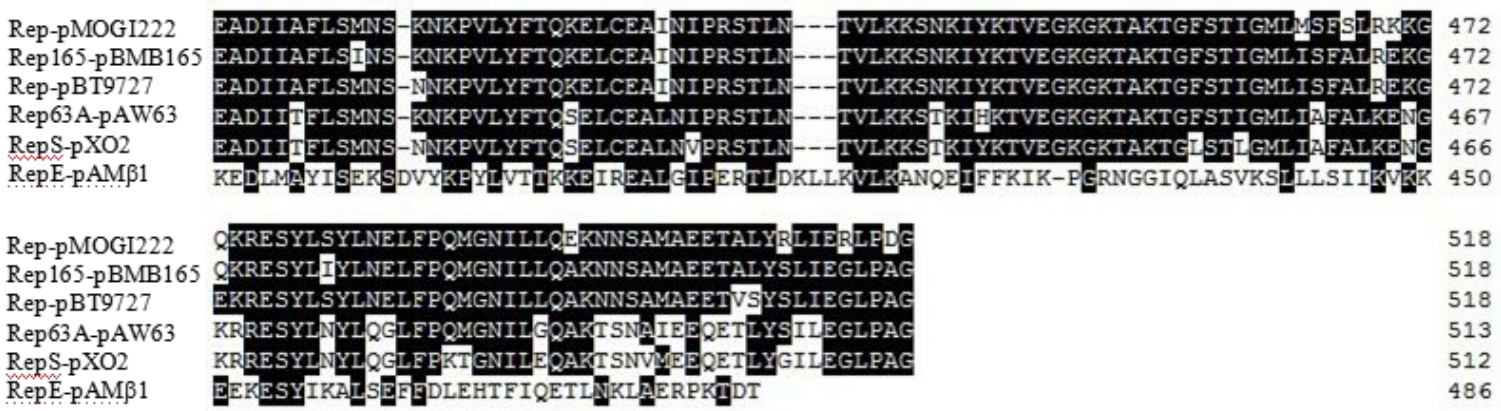


Fig. 18. Alignment of the predict Rep protein of pMOGI222, Rep165 protein of plasmid pBMB165, Rep63A protein of plasmid pAW63, RepS protein of pXO2, and RepE protein of pAMβ1 of these three plasmids. The alignment was done using the ClustalW program, and the shaded letters indicate identical amino acids.

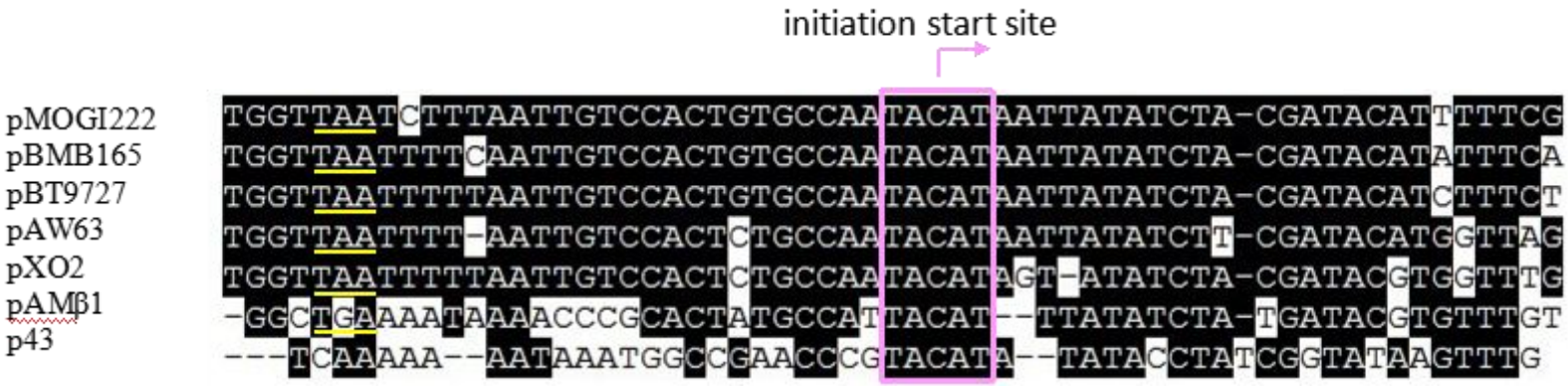


Fig. 19. Alignment of the origin of replication in pMOGI222, pBMB165, pBT9727, pAW63, pXO2, pAMβ1 and p43. The shaded letters indicate identical nucleotides. The box highlights a conserved replication initiation site 'TACAT'. The stop codons of the *rep* genes are underlined.

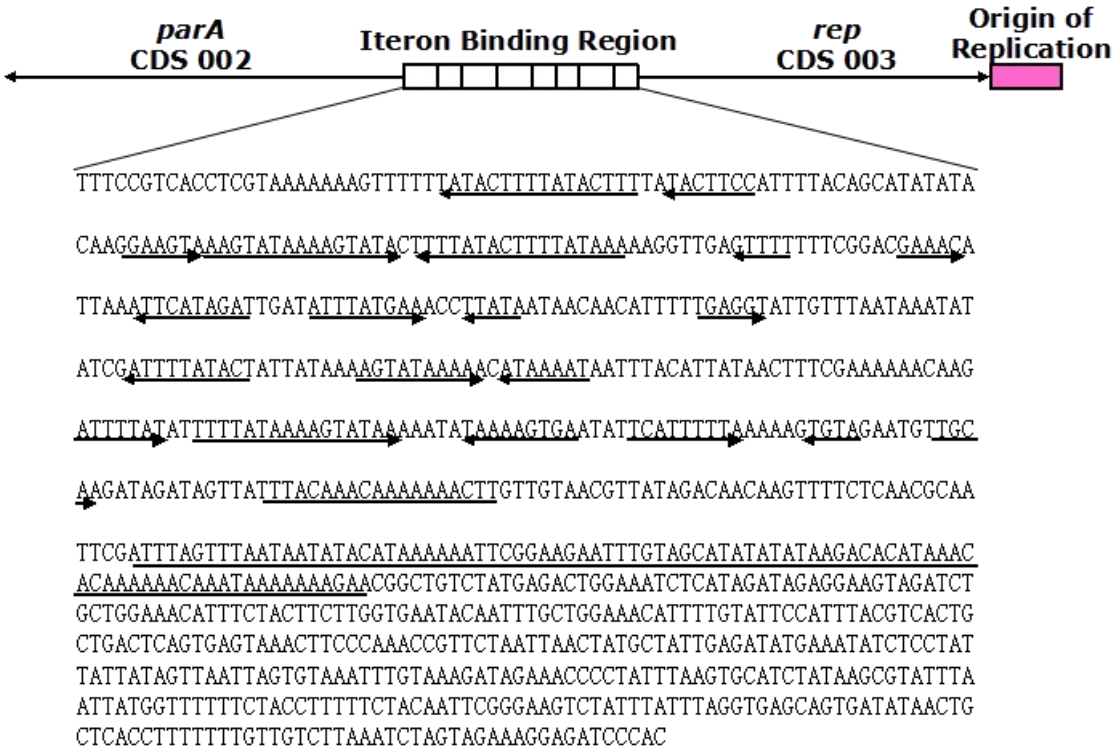


Fig. 20. The pMOGI222 replication region. The pMOGI222 iteron-binding region is between the copy number control gene *parA* and the replication gene *rep*. Arrows on the nucleotide sequence identify the iteron-binding repeat. The conserved regions of high A+T content are underlined.

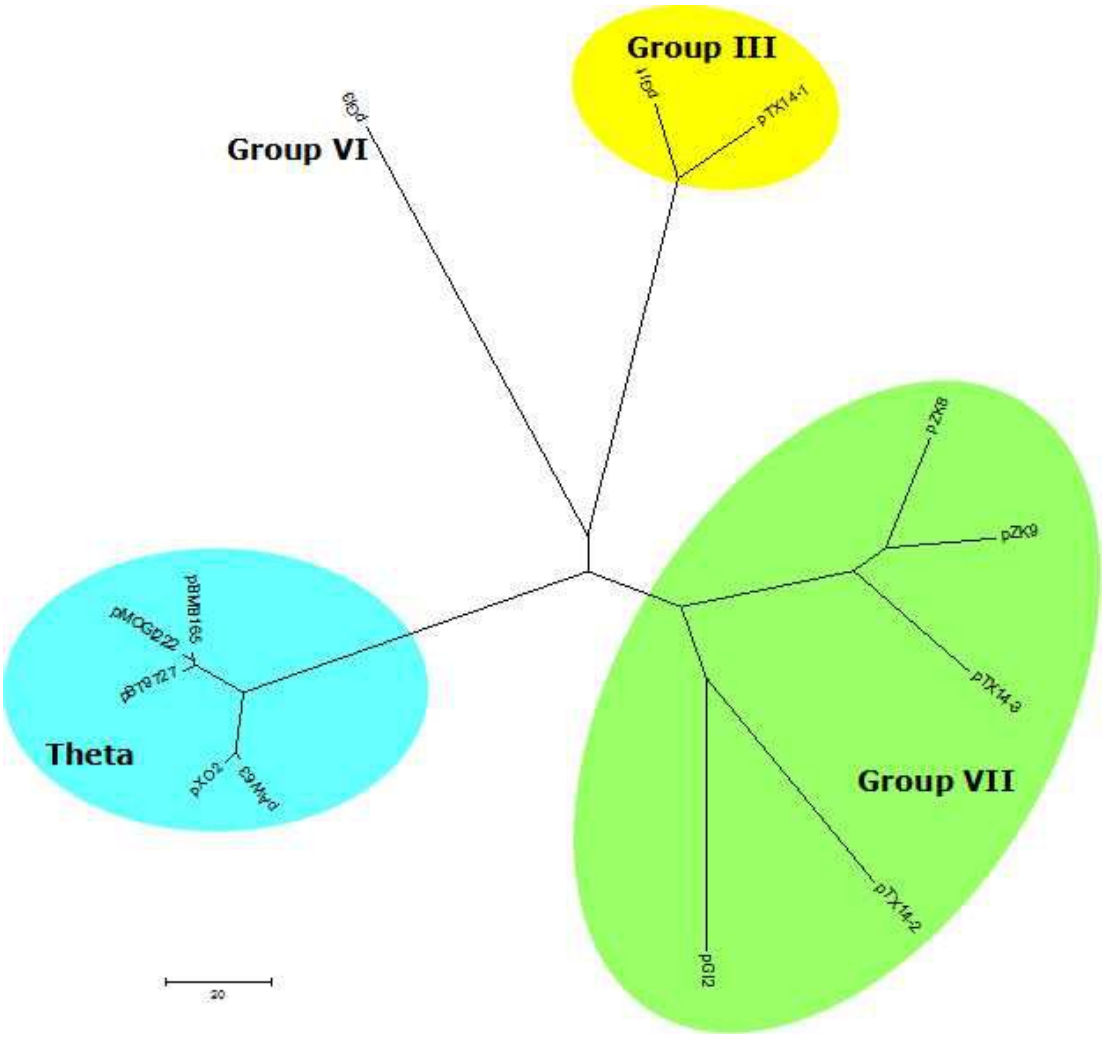


Fig. 21. Comparison and clustering of *B. cereus* group plasmids based on replication. Replication, as identified by annotation, were aligned and compared using CLUSTALW. The unrooted neighbor-joining phylogenetic trees were generated and displayed with MEGA 5.2. Theta-replicating plasmid proteins are within the blue ovals, Rolling circle replication plasmids of group VII are in the green ovals and only two members of the group III family replication proteins could be identified for inclusion.

4. Discussion

The completion of the genome sequence of *B. thuringiensis* subsp. *mogi* and comparative analysis with other bacteria species offers new information regarding the evolution of this species. Based on overall nucleotide and protein similarities, *B. thuringiensis* subsp. *mogi* is most similar to *B. cereus* G9842. The similarities including the synteny of genomes and a much similar megaplasmid, suggest that these two species are a biologically and phylogenetically divergent group whose members have developed to adapt to particular environmental conditions over evolutionary time.

The species in *B. cereus* group were classified as distinct because of the great relevance of their phenotypical differences, which formed the basis for their classification. *B. anthracis* strains are capable of capsule formation and the production of toxins that led to carbuncles in animals and in humans, causing the disease known as anthrax. *B. thuringiensis* forms a parasporal crystal that is active on larvae of a variety of insect orders. *B. cereus* lacks both of those characteristics and can cause food contamination. However, comparison of their 16S rRNA revealed less than 1% divergence between them (Ash *et al*., 1991).

In addition, many studies discovered that members of the *B. cereus* clade are very closely related in terms of phylogenic evolution as *B. thuringiensis* may resemble *B.*

cereus when losing its characteristic plasmids (Carlson *et al.*, 1994; Helgason *et al*., 1998; 2000), and vice versa, a *B. cereus* strain may display characteristic functional properties of *B. thuringiensis* or *B. anthracis* when it acquires plasmids of *B. thuringiensis* or *B. anthracis*, such as pBtoxis (Hu *et al.*, 2005) or pXO1/pXO2 (Hoffmaster *et al*., 2004).

On the other hand, despite the high identity of *B. thuringiensis mogi* and *B. cereus* G9842, there are two megaplasmids harboring several *cry* genes that are unique in *B. thuringiensis* subsp. *mogi.* The plasmid analysis of species belonging to the *B. cereus* group has become more important over the past few years due to its direct relationship with the pathogenic phenotype and the mobilization capabilities of these extrachromosomal elements (Hoffmaster *et al*., 2004; Rasko *et al*., 2005). It is crucial that the mechanisms implicated in replication, maintenance and gene transfer be understood, especially those involving toxin coding plasmids, which can affect public health. The variation of the chromosomes might be due to duplication and a complex and dynamic evolutionary process.

Chapter 3. Molecular cloning and characterization of mosquitocidal protein genes from *B. thuringiensis* subsp. *mogi*

ABSTRCT

To investigate the role of six three-domain *cry* genes (including *cry19Bb1, cry73Aa* with *cry40orf2, cry20Bb1, cry27Ab1, cry4Aa* and *cry56Ba1* with *cry39orf2*) in crystal production of *B. thuringiensis* subsp. *mogi*, the transcription level of these toxin genes were analyzed by quantitative PCR (qPCR). The results clearly indicated that all of these *cry* genes were successfully transcribed in wild type *B. thuringiensis* subsp. *mogi* strain in different expression time with different maximum levels. These *cry* genes were cloned to the *Escherichia coli-B. thuringiensis* shuttle vector, pHT1K, under the control of its own promoter, then introduced into an acrystalliferous *B. thuringiensis* Cry$^-$B strain for further molecular characterization. Another vector p1KSD, which containing a strong chimeric *cyt1Aa* promoter combined with the STAB-SD sequence was constructed and used to over-express the *cry* genes. To determine the function of the *cry39orf2* and over-express the *cry56Ba1* in *cry56Ba1* operon, different combinations of Cry56Ba1 and Cry39ORF2 were synthesized in strain Cry$^-$B. The stable inclusions in recombinant cells suggests that Cry39ORF2 assists in synthesis and crystallization of Cry56Ba1 by functioning like the C-terminal domain characteristic of Cry protein in the 130 kDa mass range. In addition, the

increased Cry56Ba1 yield under the *cyt1A-p*/STAB-SD promoter has broadened the possibility of application in other toxins.

Key words: *B. thuringiensis,* three-domain *cry* gene, quantitative PCR, over -expression.

1. INTRODUCTION

The mayor determinants of the insecticidal properties of *B. thuringiensis* bacteria are the δ-endotoxins produced during bacterial sporulation, which form two multigenic groups, *cry* and *cyt* (de Maagd *et al*., 2001). The nomenclature of Cry proteins is based on their primary sequence identity, and they have been classified in 70 subgroups. Different proteins not related phylogenetically form part of this classification, such as the group of three domain (3d-Cry) toxins, the mosquitocidal-like (Mtx-like) Cry toxins and binary-like (Bin-like) Cry toxins. The Mtx-like and the Bin-like toxins have similarity with the Mtx or Bin toxins produced by *B. Sphaericus*, although in the case of *B. Sphaericus* these toxins are toxic against mosquitoes whereas in *B. thuringiensis* they are toxic against coleopteran larvae (de Maagd *et al*., 2003).

Some similarity have been found in the 3d-Cry proteins. Domain I (a 7 α-helical bundle) is equipped for pore formation in insect epithelial membrane. Domain II (a triple β-sheet structure) may be responsible for receptor recognition. Domain III (a

β-sandwich region) may protect the toxin from further degradation during proteolytic processing or moderate toxin bilayer and toxin-toxin interactions (Honee and Visser, 1993; Vontersch *et al.*, 1994).

Among these toxins, the lineage of 3d-Cry toxins represents the largest group with more than 53 different subgroups of Cry toxins (Crickmore *et al*., 2011). One particular feature of the members of the 3d-Cry group is the presence of protoxins with two different lengths, 65 and 130 kDa. The main difference between the 65 and 130 kDa 3d-Cry toxin is the C-terminal extension that is found in the 130 kDa protoxins and is dispensable for toxicity, as it is cleaved by proteases present in the larval midgut (de Maagd *et al.*, 2001).

Since the first cloning of the *cry1Aa* gene from *B. thuringiensis* subsp. *kurstaki* HD-1 (Schnepf and Whiteley, 1981), more than 100 *B. thuringiensis* toxin genes have been cloned until 1981. The great number of sequences known to date is mostly a result of the strong interest in finding novel Cry proteins, with the focus on three main purposes: (i) the search for a new range of activities, (ii) the search for higher levels of toxicity, and (iii) the search for alternative toxins in case of resistance development (Noguera and Ibarra, 2010).

According to the genome sequence of the *mogi* described in Chapter 2, there are at least 17 toxin-related genes existed in *B. thuringiensis* subsp. *mogi,* and only the *cry* genes which possessed three domains were chose to do the further research in this

chapter. The search for novel Cry toxins has followed this strategy, construct a recombinant vector and expression it in acrystalliferous mutants of *B. thuringiensis* strain. Meanwhile, the conspicuously increase in Cry protein yield obtained with *cyt1A* promoters combined with STAB-SD sequence suggested that this combination might be useful for increasing yield of other toxins.

2. MATERIALS AND METHODS

2.1 Bacterial strains and growth media

The novel serogroup *B. thuringiensis* subsp. *mogi* (H3a3b3d) strain used in this research, was isolated from fallen leaves, sampled in a forest region of the city of Mungyeong, as previously described (chapter I) (Roh *et al*., 2009). Cloned *cry* genes and recombinant plasmid constructs were amplified in *Escherichia coli* Top10 [F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu) 7697 galE15 galK16 rpsL(Str^R) endA1 λ^-]. Cry^-B, an acrystalliferous *B. thuringiensis* subsp. *kurstaki* strain, used as the host for expression of the *cry* genes, was kindly provided by Dr. M. Ohba (Institute of Biological, Faculty of Agriculture, Kyushu University, Japan).

B. thuringiensis was grown at 28ºC with vigorous shaking in SPY medium for DNA preparation and GYS medium for expression of crystal proteins (Kronstad *et al*., 1983; Li *et al*., 2002; Nickerson and Bulla, 1974). The LB medium was used as a

primary culture of *B. thuringiensis* and in *E.coli* culture for plasmid preparation.

2.2 Preparation of RNAs for reverse transcription PCR of target *cry* genes

B. thuringiensis strain was grown in LB. Total RNAs were isolated from 2 ml portions of wild type *B. thuringiensis mogi*. From each aliquot taken during the sporulation phase, 2 ml were processed by precipitating their cell content, resuspending the pellet and washing several times by centrifugation in TE buffer (Tris 50 mM, EDTA 10 mM, pH 8.0). Last pellet was resuspended in 500 μl TES buffer (Tris base 50 mM, EDTA 10 mM, sucrose 20%, pH 7.5) supplemented with 10 mg lysozyme, and incubated at 37°C and 75 rpm for 30 min. Suspension was then pelleted at 4000 rpm for 10 min and treated with 1 ml Trizol (Invitrogen) to extract RNA according to the manufacturer's instructions. Total RNA (~1 μg) was then treated with 2 U DNAse I (Rnase-free, Takara) and evaluated by PCR amplification. Once negative PCR amplification was corroborated, specific fragments were amplified by RT-PCR when using bioneer RT-PCR premix, 150 ng total RNA, 100 ng primers (see Table 9) reaction mixture. Amplification conditions were: an initial step at 42°C for 60 min, and denaturation of 5 min at 94°C, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 57°C for 30 s, and polymerization at 72°C for 1 min, each cycle. Amplification finished with an extension step at 72°C for 5 min. Amplification products were visualized in 0.8% agarose gels.

Table 9. Nucleotide sequences of forward and reverse primers used for amplification of the target *cry* genes in reverse transcription- polymerase chain reaction (RT-PCR).

Primer	Sequence (5'-3')	Target gene
Fw-19RT	CGCTGCATGGAAACAAAATA	*cry19Bb1*
Re-19RT	CCAAAAAGGGACCATTCTCA	
Fw-73RT	CGTGATGCTAGTATGTTTGG	*cry73Aa*
Re-73RT	CATCTCGGTTCTATCTAGCC	
Fw-20RT	CCTCAAAATGGGCAACAAGT	*cry20Bb1*
Re-20RT	CGAATAGGATCATACGTTGG	
Fw-27RT	TTGGAAGAAAAACCCGAATG	*cry27Ab1*
Re-27RT	TTCCATTTGTTCTAACCGCC	
Fw-4RT	CTCAACCTTATGCAGATTGTGAT	*cry4Aa*
Re-4RT	TGCGGCTTGTGCATAAGT	
Fw-56RT	CCGGGTTATGAAGCCTTAC	*cry56Ba1*
Re-56RT	GTCCTTCAAGTGGCTCTCC	

2.3 Analysis of mRNA expression by qPCR

For monitoring the expression levels of different *cry* genes in *B. thuringiensis mogi*, qPCRs were conducted as follow. Total RNAs (1 μg) were reverse transcribed to cDNA with the QuantiTect® Reverse Transcription Kit (Qiagen) according to the manufacturer's instructions. For qPCRs, primers specific to *cry19Bb1, cry73Aa, cry20Bb1, cry27Ab1, cry4Aa, cry56Ba1,* and 16S rRNA were designed (Table 10). The 16S rRNA qPCRs were performed as internal controls to normalize the amount of RNA input. qPCR was carried out using the Bio-Rad CFX Manager real-time detection system (Bio-Rad). Reaction mixtures (20 μl) contained 10 μl 2×EvaGreen qPCR mastermix (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 0.2 mM each dNTP, 6 mM $MgCl_2$), 4 pM each primer and 25 ng of diluted cDNA. The thermal cycling conditions were denaturation at 95°C for 5 min, followed by 40 cycles of 95°C for 30 s, 58°C for 30 s, and 72°C for 30 s. This was followed by a melting curve program of 55 to 95°C with a heating rate of 0.5°C and final cooling at 10°C. The data obtained from qPCR were analyzed by relative quantification using the ΔC_T (cyclic threshold) method. One-way analysis of variance (ANOVA) was used to compare the mean differences of relative quantification in different strains at each time point with a significance level of $P < 0.05$. Samples were collected from three independent experiments.

Table 10. Nucleotide sequences of primers used for amplification of the target *cry* genes in qPCR.

Primer	Sequence (5'-3')	Target gene
Fw-16SrRNA	CGTAGGTGGCAAGCGTTATCC	*16S rRNA*
Re-16SrRNA	TCCTCTTCTGCACTCAAGTCTCC	
Fw-19Q [a]	ATCCGTTTACATCTATCCCCATCTCAC	*cry19Bb1*
Re-19Q	CCGCTCAAGCATACCACATTCG	
Fw-73Q	CGCCGAGTCGCATACTATGATTTC	*cry73Aa*
Re-73Q	GTCTCCCTGTCCTACGCCAAAG	
Fw-20Q	ACAGTCTGGATCTACCACACCTTG	*cry20Bb1*
Re-20Q	AACTCCGCCATAACTTCGTTGTATAG	
Fw-27Q	TGGTAACACGCTACGGTAAGGAG	*cry27Ab1*
Re-27Q	GGTTCATCACAGACTGGGACAATG	
Fw-4Q	CTCAACCTTATGCAGATTGTGAT	*cry4AA*
Re-4Q	GCTAGAACTGGCGCTGCTAT	
Fw-56Q	GCTCCAATTACTGGCGGAACATC	*cry56Ba1*
Re-56Q	ACTCTCAGCATTAGCGGGAAAGG	

[a] The superscript letters Q indicate primers for quantitative real-time PCR.

2.4 PCR amplification *cry* genes, cloning and sequence analysis

All PCR amplifications were performed with the DNA Thermal Cycler (Perkin Elmer Co., USA). For amplification of the active region of *B. thuringiensis*, the PCR was performed using an Expand Long Template PCR system (Roche Diagnostics GmbH, Germany) and a GeneAmp 2400 PCR system thermocycler (Perkin Elmer, Boston, MA) for 33 cycles. For PCR, 0.1 μg of total DNA from *B. thuringiensis mogi* was mixed with a solution containing each primer (Table 11) at a concentration of 20 pM, each dNTP at a concentration of 10 mM, and 1.25 unit of PrimeSTAR GXL DNA polymerase in PCR buffer (Takara, Japan). Conditions were as follows: 94°C for 1 min, 55°C for 1 min, 68°C for 4 min, and then 7 min of termination at 68°C. Then the PCR products were cloned into the pHT1K plasmid vector and sequenced (Fig. 22).

The expression vector (pHT1K, contained an erythromycin resistance gene marker) used in this study was designed to express the different *cry* genes in *Bacillus* spp. This vector encoded the insecticidal *cry* genes of *B. thuringiensis* subsp. *mogi* under the control of its endogenous promoter in a minimal *E. coli-B. thuringiensis* shuttle vector. All of these recombinant plasmids were transformed into *E. coli* Top10 and confirmed by nucleotide sequencing.

2.5 Transformation of *B. thuringiensis*

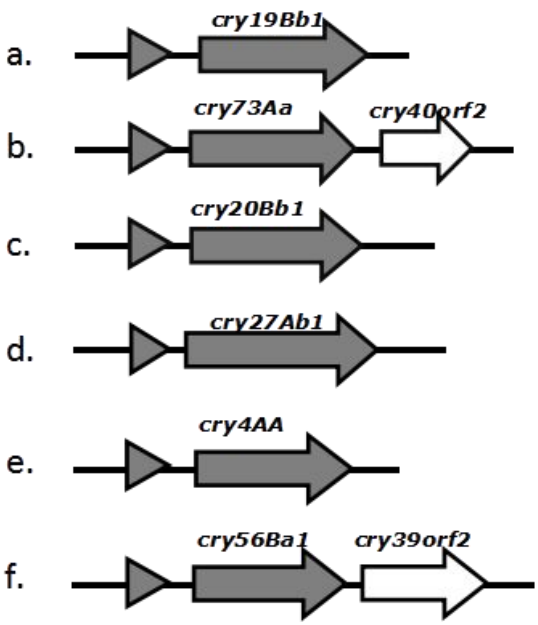
Once confirmed the sequences, these new constructs were used to transform the acrystalliferous *B. thuringiensis* subsp. *kurstaki* strain Cry$^-$B by electroporation following the protocol (25 μF, 2 kV, 400 Ω) commonly used for the transformation of *B. thuringiensis* (Lereclus *et al*, 1989) in 0.2 cm electrode gap electroporation cuvettes (Bio-Rad, USA). After the pulse, the electroporated mixture was added to 1 ml of brain heart infusion (BHI) (Beckton, Dickinson and Company, Sparks, MD) and incubated with gentle shaking (60 rpm) for 1 h at 30°C. Transformants were selected on nutrient agar plates containing erythromycin (25 μg/ml).

Table 11. Nucleotide sequences of primers used for amplification of the target *cry* genes to pHT1K vector in PCR.

Primer [a]	**Sequence (5'-3')**	**Target gene**
Fw-19	ACGAATTCGAGCTCGCATATTACAGCTTTTCCAC	*19Bb1*
Re-19	CCAGTGCCAAGCTTGCAGTTCACTCTTTTCTG	
Fw-73	CCATGTGCAGTTGTTGTATC	*73Aa+40orf2*
Re-40	CAAGGTTCGTCGACAATTC	
Fw-20	GTATACTGAGCTCTTGATTTCG	*20Bb1*
Re-20	GAATAGAGGCCATGCATGC	
Fw-27	CAAATCATGGTACCCAGCTAC	*27Ab1*
Re-27	GACATAGAACGTCGACACTTGG	
Fw-4	ACGAATTCGAGCTCGGAATCGCGCATATTTCAGG	*4Aa*
Re-4	CCAGTGCCAAGCTTGGAGCCAACTAGATATG	
Fw-56	GATAGAGCTCCATTGGTTATTGAAC	*56Ba+39orf2*
Re-39	CGTGTCGACAGATACCTCGTCAG	

[a] Primers for cloning target genes to pHT1K vector.

A



B

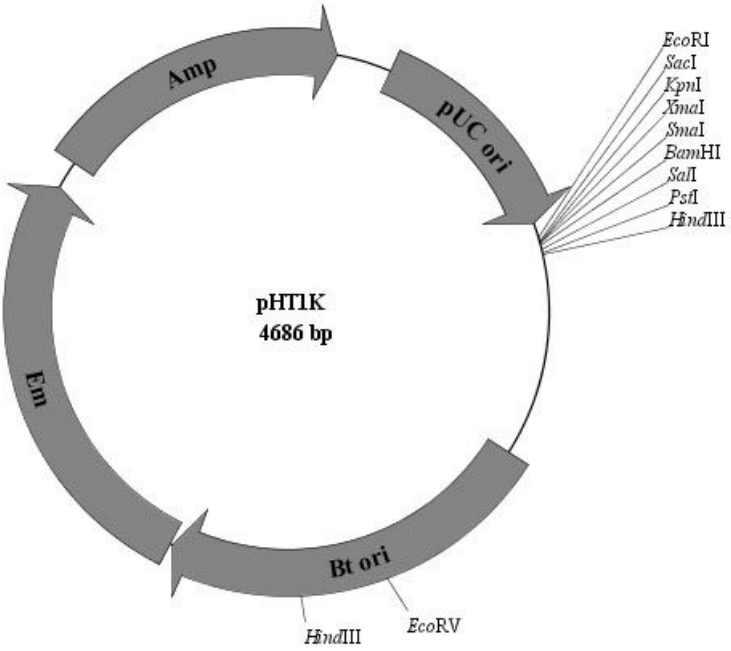


Fig. 22. Cloning of mosquitocidal *cry* genes from *B. thuringiensis* subsp. *mogi*. (A) The arrangement of six *cry* genes from *mogi.* (B) The *E. coli-B. thuringiensis* shuttle vector, pHT1K. Amp, amplicillin resistant gene; Em, erythromycin-resistant gene;

pUCori, *E. coli* Replication origin, Bt ori, *B. thuringiensis* replication origin.

2.6 Construction of recombinant plasmids for over expression of Cry56Ba1 and functional analysis of Cry39ORF2

The primers used to amplify different regions of the *cry56Ba1* operon are listed in Table 12, and the plasmids constructed using the PCR products are illustrated in Fig. 23. A series of recombinant plasmids were constructed for two different purposes:

(i) Analysis of the *cry56Ba1* operon

To determine the function of Cry39ORF2 encoded by the second gene in the *cry56Ba1* operon (Fig. 22A, panel f), DNA fragments containing *cry56Ba1* and *cry39orf2* were obtained using PCR, with *mogi* DNA as the template. The primer sets Fw-56 + Re-39 and Fw-56 + Re-56 (Table 12) were used to amplification fragments *cry56Ba1* + *cry39orf2* (Fig. 23, panel 1) and *cry56Ba1* only (Fig. 23, panel 2). Both of these products were cloned to pHT1K vector (Fig. 22B) and expressed under their original promoters.

(ii) Improvement of Cry56Ba1 synthesis, crystal topology, and determination of Cry39ORF2 function

To improve the synthesis of Cry56Ba1, plasmids containing the following elements were constructed as follows: fragment *cry56Ba1*+*cry39orf2* (Fig. 23, panel 3); *cry56Ba1* (Fig. 23, panel 4) and *cry39orf2* only (Fig. 23, panel 5). The primer sets which used for amplification the different constructs were Fw-56inf + Re-39inf, Fw-56inf + Re-56inf and Fw-39inf + Re-39inf, respectively (Table 12). All of these

products were cloned to p1KSD (Fig. 25) vector and expressed under the strong chimeric *cyt1A*-p/STAB-SD (Fig. 24) promoter.

All of these recombinant plasmids were transformed into *E. coli* Top10 and the integrity of these sequences was confirmed by restriction enzyme digestion and sequencing analysis. These new construct were then electroporated into *B. thuringiensis* Cry^-B as described before.

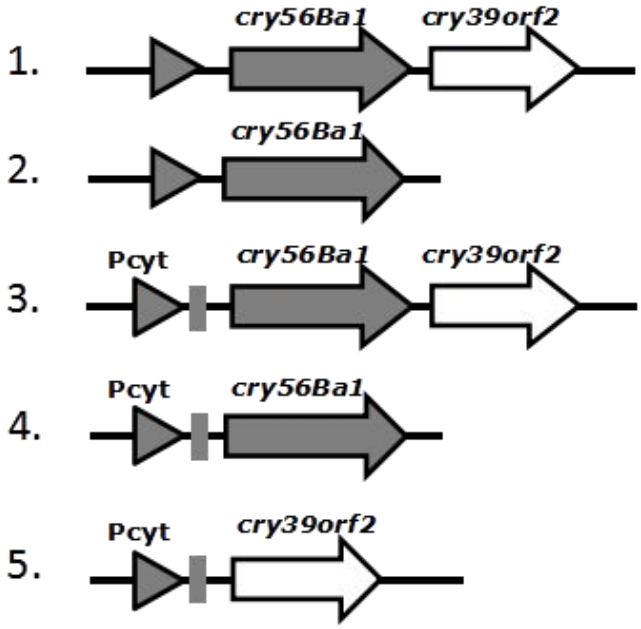


Fig. 23. Schematic illustration of the different constructs containing *cry56Ba1* or *cry39orf2* or both. Construction of No. 1 and 2, expression were under the control of the original wild-type promoter of *cry56Ba1*. Construction of No. 3, 4 and 5, expression were under the control of the c*yt1A*-p/STAB-SD sequence.

Table 12. Primers used for amplification of constructs of *cry56Ba1* and *cry39orf2*.

Primer	Sequence (5'-3')
Fw-56[a]	GATAGAGCTCCATTGGTTATTGAAC
Re-56[a]	CCAAGCTTGCATGCATAC
Re-39[a]	CGTGTCGACAGATACCTCGTCAG
Fw-56inf[b]	GAAGGCTTTTTCTAGTAAGGACTACATAAGGAGTG
Fw-39inf[b]	GAAGGCTTTTTCTAGTAGGTAGTAATCCTGTTC
Re-56inf[b]	GGCCAAGCTTGGATCTACCTACTTCATTACATAC
Re-39inf[b]	GGCCAAGCTTGGATCGCAATGATTTTAATTC

[a] Primers for cloning target genes to pHT1K vector

[b] Primers for cloning target genes to p1KSD vector.

Fig. 24. Upstream region and schematic illustration of constructs used to synthesize Cry proteins. A: Nucleotide sequence of the 660 bp fragment containing *cyt1A* promoters combined with STAB-SD sequence. Sigma E-like and sigma K-like promoters of *cyt1A* are shown, respectively, in boxes or underlined. The STAB-SD sequence is highlighted as a black box. B: Nucleotide sequence of the terminal sequence. Arrows indicate orientation of inverted repeat sites.

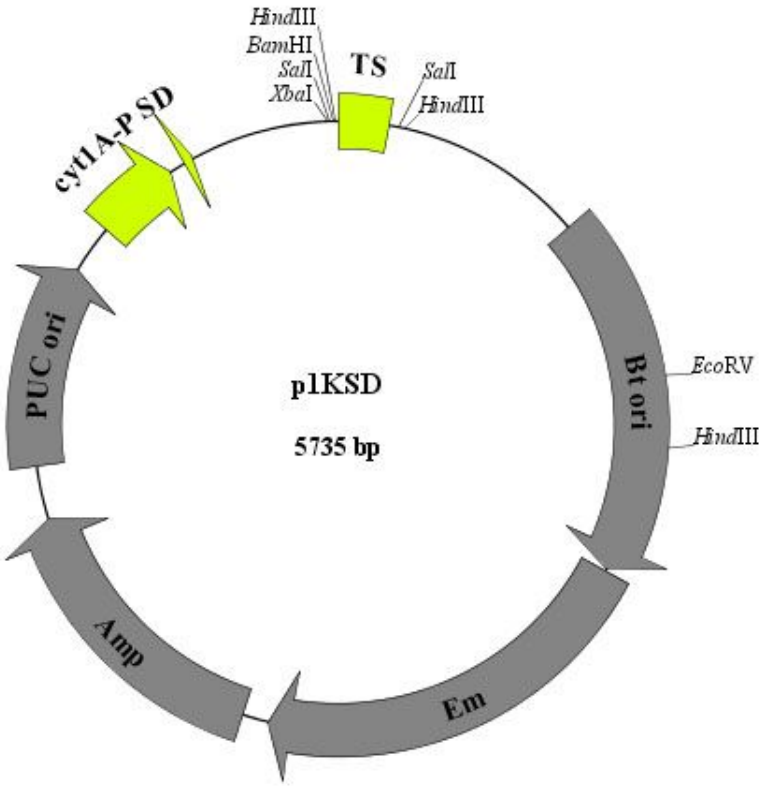


Fig. 25. The over-expression vector p1KSD. Amp, amplicillin resistant gene; Em, erythromycin-resistant gene; pUCori, *E. coli* Replication origin, Bt *ori*, *B. thuringiensis* replication origin.

2.7 Morphological observation and SDS-PAGE

Parasporal inclusions were purified by the method of Thomas and Ellar (1983). Crystal morphology of the isolates were examined by phase-contrast microscopy or transmission electron microscopy as described in Chapter 1 (see materials and methods 2.5). For SDS-PAGE samples, cells were cultured on NAE (nutrient agar plates containing erythromycin 25 μg/ml) medium plate at 28ºC and harvested after autolysis. SDS-PAGE was performed on a 12% separating gel with 5% stacking gel. The gel was stained with 0.1% Coomassie brilliant blue (Sigma Co., St Louis, MO, USA).

2.8 N-terminal sequencing

The N-terminal sequence of Cry20Bb1 and Cry56Ba1 proteins from recombinant *B. thuringiensis* Cry$^-$B strains were determined as follows. The inclusions from recombinant Cry$^-$B strains (CB/pHT1K-20Bb1 and CB/p1KSD-56Ba1+39orf2) were harvested and washed 3 times with a washing solution (0.5 M NaCl, 2 mM EDTA), then separated on a 10% of SDS-PAGE gel. The gel was transferred to a PVDF membrane (Amershem Hybond^{TM}-P, GE Healthcare). The membrane was stained with Coomassie brilliant blue R250 and the protein bands were excised and subjected to procise 491 HT protein sequencer (Applied Biosystems, USA) sequencing at the University of Korea, Korea.

2.9 Insects and toxicity assays

The mosquito larvicidal activities were assayed on 4th instar larvae of *Culex pipiens molestus* and *Culex pipiens pallens* (Diptera: Culicidae) which were grown in a container (35×25×3 cm) at 25ºC. Freeze-dried *B. thuringiensis* spores-crystal complex were suspended in double-distilled water. Suspensions were diluted to 6 or 7 different concentrations in cups in a final volume of 100 ml. Bioassays were replicated three times using 30 4th *Culex pipiens* instars of per concentration. After 48 h of exposure at 25°C, dead larvae were counted. Statistical analysis of data and 50% lethal concentrations (LC_{50}) were performed with probit analysis (Russell *et al*., 1977).

2.10 Amino acid sequence alignments

Homology searches with *cry* genes and orf2 were performed using the Basic Local Alignment Search Tool (BLAST) (BLASTP version 2.2.2; http: //www.ncbi.nlm.nih.gov/). ORF2 and Cry protein sequences that showed the highest level of identity were then aligned using MegAlign software (DNAstar, http://www.dnastar.com).

2.11 Nucleotide sequence accession number.

The nucleotide sequence data shown below (Fig. 26 - Fig. 31) are available in the GenBank, and the accession number listed in table 14.

3. RESULTS

3.1 Summary of toxin-related genes in *B. thuringiensis* subsp. *mogi*

According the CDS information from pMOGI364 and pMOGI222 (Chapter 2, table S1 and S2), there are at least 17 toxin-related genes existed in *B. thuringiensis* subsp. *mogi* (summary in Table 13). These two megaplasmids carry several insertion sequences and encode two further proteins (for example: CDS pMOGI364_340, which encodes a *cryBP1* family protein and shows homologue with P19; CDS pMOGI222_132, another 19 kDa accessory protein in *B. thuringiensis* subsp. *mogi*) with roles in promoting crystal formation and enhancing cell viability, probably by acting as chaperones (Dervyn *et al.*, 1995; Manasherob *et al*., 2001; Wu and Federici, 1993).

Six *cry* genes which contained intact 5 conserved blocks (including *cry19Bb1, cry73Aa* with *cry40orf2, cry20Bb1, cry27Ab1, cry4Aa* and *cry56Ba1* with *cry39orf2*) were chose for qPCR test and further studies. The detailed information about the nucleotide sequences and deduced amino acid sequence were shown in Fig. 26 to Fig. 31. These sequences were submitted to GenBank, and the accession numbers are given in Table 14.

As shown in Fig. 27, sequence of *cry73Aa* operon include two open reading frames oriented in the same direction and separated by 64 bp were identified (Fig. 27): *cry73Aa* (2,010 bp) and *cry40orf2* (1,494 bp). A putative ribosome binding site (RBS),

GGAGT, was identified 7 nucleotides upstream from the start codon of the *cry73Aa* gene (Fig. 27), and a sequence, AAAGGTTGTG, that could act as a RBS was identified 6 nucleotides upstream from *cry40orf2* (Fig. 27). The similar gene structure also emerge in *cry56Ba1* operon (Fig. 31), two open reading frames oriented in the same direction and separated by a 68 untranslated bp, two putative RBS sited were located 7 and 6 bp upstream of the star codon of *cry56Ba1* and *cry39orf2*, respectively.

Table 13. Summary of toxin-related genes in pMOGI364 and pMOGI222.

Locus_tag	Size (aa)	Strand	Annotation
pMOGI364_238	684	+	pesticidal crystal protein cry19Aa
pMOGI364_261	497	-	Cry8Ka2 delta-endotoxin
pMOGI364_262	669	-	Cry1-like delta-endotoxin
pMOGI364_287	340	-	mosquitocidal toxin gene
pMOGI364_322	299	-	35.8-kilodalton mosquitocidal toxin
pMOGI364_328	256	+	mosquitocidal toxin gene
pMOGI364_340	190	+	cryBP1 family protein
pMOGI364_348	722	+	Cry20-like delta endotoxin
pMOGI364_356	194	-	Cry20-like delta endotoxin
pMOGI222_106	220	-	mosquitocidal toxin
pMOGI222_108	825	+	Pesticidal crystal protein cry27Aa
pMOGI222_132	185	+	19kda accessory protein
pMOGI222_133	683	+	pesticidial crystal protein cry4AA
pMOGI222_134	506	-	crystal protein ET69
pMOGI222_186	557	-	mosquitocidal toxin protein
pMOGI222_235	562	-	Cry39ORF2 protein
pMOGI222_236	659	-	Cry56Aa-like protein

Table 14. Amino acid sequences of putative crystal proteins of *B. thuringiensis* subsp. *mogi* strain.

Protein No.	Locus_tag[a]	Protein Name	Predicted size	Protein size	GenBank accession No.
1	pMOGI364_238	Cry19Bb1	683 aa	78 kDa	KC182376
2	pMOGI364_262	Cry73Aa	669 aa	77 kDa	KC182375
	pMOGI364_261	Cry40ORF2	497 aa	57 kDa	KC182377
3	pMOGI364_348	Cry20Bb1	753 aa	83 kDa	KC182372
4	pMOGI222_108	Cry27Ab1	825 aa	94 kDa	KC182373
5	pMOGI222_133	Cry4Aa	683 aa	77 kDa	
6	pMOGI222_236	Cry56Ba1	659 aa	73 kDa	KC182374
	pMOGI222_235	Cry39ORF2	562 aa	64 kDa	KC182378

[a] locus tag means the *cry* genes located site.

Translation map – *cry19Bb1*

```
                                                                  -35
CATATTACAGCTTTTCCACTCCACAATCCACATGTGGATTACGATTATCCAGCTACTATA      60
                  -10
TCTATAAAATTTTGTATAATGAGATTAATTAAAAGATGTGTAAAGGCTGAATAATCCCCT     120
GGATACAAACATACGTATAAGATTGAAGAATAGTATCGATTCTGTCAATTACTTAGATAG     180
                          RBS                M  H  P  Y  Q  N  K  N  E
ATACTTCTTTAATATGAGCACGGGGGGAACTAAGATGCATCCTTATCAAAATAAGAATGA     240
Y  E  I  L  D  A  T  Q  N  N  C  H  M  S  N  C  Y  P  K  Y
ATATGAAATCTTAGATGCTACACAAAATAATTGTCACATGTCTAATTGTTATCCCAAGTA     300
P  L  A  N  D  P  Q  M  Y  L  R  N  T  H  Y  K  D  W  I  N
CCCACTAGCAAATGATCCTCAAATGTATTTGCGCAACACCCATTATAAGGATTGGATAAA     360
M  C  E  E  A  S  Y  A  S  S  G  P  S  Q  L  L  K  V  G  G
TATGTGCGAGGAAGCTTCTTATGCATCTTCAGGTCCTTCACAATTACTTAAAGTTGGAGG     420
S  I  V  A  K  I  L  G  M  I  P  E  V  G  P  L  L  S  W  M
TTCTATAGTTGCTAAAATTCTTGGAATGATTCCTGAAGTTGGTCCTCTTTTAAGTTGGAT     480
V  S  L  F  W  P  T  I  Q  E  K  N  T  V  W  E  D  M  I  K
GGTATCTTTATTTTGGCCAACTATTCAAGAAAAAAATACTGTTTGGGAAGATATGATAAA     540
Y  V  A  N  L  L  K  Q  E  L  T  N  Y  T  L  N  R  A  T  S
GTATGTAGCAAATCTGTTAAAACAAGAATTAACAAATTATACACTTAACCGTGCCACAAG     600
N  L  F  G  L  N  E  S  L  N  I  Y  N  R  A  L  A  A  W  K
TAATTTATTTGGATTAAATGAATCTTTGAACATATATAACCGAGCCCTCGCTGCATGGAA     660
Q  N  K  N  N  F  A  S  G  E  L  V  R  A  Y  I  N  D  L  H
ACAAAATAAAAACAATTTCGCAAGTGGGGAACTTGTAAGGGCATATATAAATGATCTTCA     720
I  L  F  T  R  D  I  Q  S  D  F  S  L  G  G  Y  E  A  V  L
TATACTCTTTACAAGAGATATTCAATCAGATTTCTCATTAGGAGGCTATGAAGCCGTATT     780
L  P  S  Y  A  S  A  A  N  L  H  L  L  L  L  R  D  V  A  I
ATTACCTTCATATGCAAGTGCTGCCAATCTTCATTTACTATTGTTACGTGATGTTGCAAT     840
Y  G  K  E  L  G  Y  P  L  E  D  V  E  F  Y  Y  N  E  Q  K
TTACGGAAAAGAATTAGGATATCCCTTAGAGGACGTAGAATTTTATTATAATGAGCAAAA     900
F  Y  T  E  K  Y  S  N  Y  C  V  N  T  Y  K  A  G  L  E  L
GTTCTATACAGAAAAATATAGTAATTATTGTGTAAATACGTACAAAGCGGGTTTAGAATT     960
A  K  Q  I  G  W  S  D  F  N  R  Y  R  R  E  M  T  L  S  A
AGCAAAACAAATAGGATGGTCAGATTTTAATCGTTATCGCAGAGAAATGACTTTATCCGC    1020
L  D  I  V  A  L  F  P  L  Y  D  T  R  L  Y  P  S  K  D  G
ATTAGATATAGTTGCCTTATTCCCACTGTATGATACAAGACTGTATCCGAGTAAAGATGG    1080
K  I  H  V  K  S  E  L  T  R  E  I  Y  S  D  V  I  N  A  H
TAAGATACATGTTAAATCTGAACTAACGAGAGAAATTTACTCTGATGTTATTAATGCTCA    1140
```

V D L V L K E D K A Y F T Q V E A L Y T
TGTAGACTTAGTCTTAAAAGAAGATAAGGCATATTTTACGCAAGTTGAAGCGCTTTATAC 1200
R R P H L F T W L R G F R F V T N S I S
ACGTCGACCACATTTATTTACTTGGTTACGAGGATTTAGATTTGTAACCAATTCCATTTC 1260
S W T F L S G A Q N K Y S Y T S S S S I
TTCTTGGACATTCTTATCTGGCGCGCAAAATAAATATTCTTATACATCCTCTAGTTCAAT 1320
E N G P F L G Q D T E Y G G T S S N M V
TGAGAATGGTCCCTTTTTGGGTCAGGATACAGAATATGGTGGAACCTCTTCTAATATGGT 1380
I P E N Q Y I Y N L W T K N Y E W I Y P
CATTCCAGAAAATCAATATATTTATAATTTATGGACCAAAAATTATGAATGGATTTACCC 1440
W T D P V N I T K I N F S L T D N N S S
TTGGACTGATCCAGTAAATATTACAAAAATTAATTTTTCTCTAACAGATAATAATTCCTC 1500
N E V I Y G A E R I N K P T V R T D F N
TAACGAAGTTATCTATGGTGCAGAAAGAATTAATAAACCTACTGTTCGAACAGATTTCAA 1560
F L L N K E G T G P A T Y Y D Y N H I L
TTTTCTGCTTAACAAAGAAGGCACTGGTCCTGCGACATATTATGATTATAATCACATTTT 1620
S Y T L I N G S T A G Q K R H G Y S F A
ATCATATACACTAATAAACGGAAGTACTGCTGGTCAAAAAAGGCACGGATATTCATTTGC 1680
F T H S S V D P Y N K I A T D K I T Q I
TTTTACaCATAGTAGTGTGGACCCATATAACAAAATTGCCACAGATAAAATTACTCAAAT 1740
P A V K S N G W M F F G D V L K G P G H
TCCTGCGGTGAAAAGTAATGGATGGATGTTTTTTGGTGACGTATTAAAAGGTCCTGGCCA 1800
T G G D L V T L S N G G R Y T L N I I F
TACAGGTGGAGATTTAGTGACTCTTAGTAATGGGGGTAGATATACACTAAATATTATTTT 1860
P A Q A Y H I R I R Y A S N G D G E M G
CCCCGCTCAAGCATACCACATTCGTATTCGGTATGCTTCTAATGGTGACGGTGAGATGGG 1920
I D V N G S G Y T R F S I K S T F S H N
GATAGATGTAAACGGATCGGGGTATACCCGTTTTAGTATAAAGAGCACTTTTTCTCATAA 1980
N Y N D L N F Q D F N L M D T S F I Y N
TAATTATAATGATTTAAACTTCCAAGATTTCAATTTAATGGACACATCTTTTATTTACAA 2040
A T Y T G S Q T I W L Y S Y A T A R V I
TGCAACTTATACAGGATCACAGACTATATGGTTATACAGTTATGCAACAGCACGGGTAAT 2100
I D K I E F I P V G T F A N Q L L E E T
TATAGATAAAATTGAATTTATACCAGTTGGTACTTTTGCAAATCAATTATTAGAAGAAAC 2160
Q C Y N Y N Q N M D N T Y Q P S Y A N T
ACAATGTTATAACTATAATCAAAACATGGATAATACATACCAACCAAGCTATGCCAATAC 2220
Y N H N S S N M H N Q S Y N N *
CTACAATCACAACTCAAGTAATATGCATAATCAAAGTTATAACAATAATTATAACCAAAA 2280
IR IR
CATGGATAATACATACCAACCAAGCTATGACAATACCTACAATCACAACTCAAGTAATAT 2340

GCATAATCAAAGTTATAACAATAATTATAACCAAAACATGGATAGCATGTACAACAATAA 2400
CTATAGCCAAAGTACTAATGATATGTACCCTCAAGAATATACTAACAGCAATAACCAAAA 2460
ATTCGGCTGTACATGTAATCAAGGGTATAATAACTATCCAAAATAAGTACAGAAAAGAGT 2520
GAACTGCACCCCAATTGTTAGACACAGTTTAACAATTGGAGGTGCAGTTTTTCTATGGCT 2580
AAATTTACAGCTGATGAAAAAATACAAATCGTTCTACGTTATTTGAACGGAAATGAAAGT 2640
TATCGAGAAATGGGTAGATCGCTCGGTATAAGTGACACAATCATTTTGAATTGTGTAAAA 2700
CAGTGCTCAACAATTGCAAATTCTTTTATGTAGAGGAACTCCGACCTCAAAACCCCAAAG 2760
AATTTTTCTATCACAGCGTTGTCGTAACATTTTCCTTTTCGAGACATACTCTGGACGATA 2820
GCTCTAGATTCAAGTGT 2837

Fig. 26. Nucleotide sequence and deduced amino acid sequence of the *cry19Bb1* gene. The gene is 2,052 bp in length and codes for a polypeptide of 683 amino acids. The potential -35 and -10 boxes and a putative ribosome-binding site (RBS) are overlined. The stop codon is marked with asterisks. Five conserved sequence blocks (blocks 1 to 5) are shadowed. Terminal inverted repeats (IR) are indicated below the arrow.

Translation map – *cry73Aa+cry40orf2*

```
CGGCAATCTCAATTAAATCCTTGGTAGAACGTGAAAGTCGATTCAGGTATGTAACTGCTA    60
CTGTATCTCCATTCCGCAGAAACTTTTACATTCGATCTAGCTCCAGACGATCTCGTTTGG    120
TTCCTGTCGGTCTAAATTTGGATCAGTGGTAGATACCAGTATATATCCAAACTTCATATG    180
                    -35                 -10
ATTCCTTCTTTTCAGTTTTATTGTATCAAAAACGATATAAAATTCGTATTGTAGCAAATT    240
AATTTTGTTACAGTACGATTGCTAGTTTACTTAATTTCAGCTCTTGATACCTATCCTTGT    300
CACAAAAACGGAAATTCTTTGTACATAGGTAAAATTCTCGTGATTCGGTTTTCTCTAGAA    360
AAAATGAACTGAATCCAATACATGTTCAGATTTATTTTCTACAATTTCTTTCAATGATGA    420
ATAATTAAAAATCTACAACAATCTATTCAATATTCCTCTCTTTTTTCGAGGGATAGGAAA    480
AAACATCCAAGTGTGAATTTTGTACATATAAAGGTGAATAACTCTTTCCACATTCTAAAA    540
AAACAACAAAGAAAAAATCGTTCTACAGAAATCTGAAGCTTTTAAAAAAATACATACAAT    600
ACATAAAGAGAAGGTTTAAAAAAATAGATGCCTCACAAAATATAATGGGTTTATTTGTAG    660
AAACATCGTTAGAGGAATACATTGGGATGCTGCGAATATATAGAAAGACATCTAGTATAT    720
                                   RBS       M  N  S  Y  Q
ATTCATTAGGTATCTTAATATAAGGATTACATAAGGAGTGAAAATATGAATTCATATCAA    780
N  K  N  K  N  K  Y  E  M  L  D  T  S  R  K  S  S  N  M  S
AATAAAAATAAAAATAAATATGAAATGTTAGATACTTCAAGAAAAAGCTCTAATATGTCT    840
T  C  Y  P  R  Y  P  I  A  K  N  P  Q  K  T  M  Q  N  T  N
ACTTGTTATCCTCGGTACCCAATAGCAAAAAATCCACAAAAAACCATGCAAAATACGAAT    900
Y  K  D  W  I  N  M  C  T  S  K  N  L  E  D  G  I  Y  S  T
TATAAAGACTGGATAAATATGTGTACATCAAAAAATCTTGAAGATGGCATTTACTCTACA    960
S  A  K  D  V  I  T  N  S  I  N  I  S  S  Y  I  I  S  M  L
AGTGCAAAAGATGTAATTACAAATTCTATTAATATTTCTAGTTATATAATAAGTATGTTA   1020
G  M  P  Y  L  S  S  I  V  A  I  W  G  V  L  F  N  A  L  W
GGCATGCCTTACTTATCAAGCATCGTAGCTATATGGGGGGTGCTTTTTAATGCATTATGG   1080
P  S  S  D  N  Q  W  E  P  Y  M  N  H  V  Q  G  L  I  R  R
CCTAGTTCAGATAATCAATGGGAACCTTATATGAATCATGTACAAGGTCTTATTAGGCGA   1140
E  L  Q  T  F  A  R  E  Q  A  L  R  Q  L  E  G  L  G  G  N
GAATTACAGACTTTTGCAAGAGAACAAGCACTTAGACAATTAGAGGGGTTAGGTGGAAAT   1200
L  D  L  Y  K  E  A  L  E  E  W  E  Q  D  R  D  N  Q  T  T
TTAGATTTATATAAAGAGGCATTGGAAGAATGGGAGCAAGACCGTGACAATCAAACAACT   1260
K  E  R  V  R  D  R  F  R  I  L  D  G  F  F  T  Q  Y  I  P
AAAGAAAGGGTAAGAGATCGATTCCGCATATTGGATGGTTTTTTCACGCAATACATTCCG   1320
V  F  R  I  Q  G  Y  E  V  Q  L  L  S  V  Y  T  K  V  A  N
GTTTTTAGAATCCAAGGGTATGAAGTTCAATTATTATCTGTCTATACAAAAGTTGCGAAT   1380
L  H  L  L  L  L  R  D  A  S  M  F  G  A  D  W  G  M  S  Q
CTCCATTTGCTTTTATTACGTGATGCTAGTATGTTTGGGGCCGATTGGGGAATGAGCCAA   1440
T  N  I  N  D  N  Y  N  R  Q  M  N  L  T  S  L  Y  T  N  H
```

ACTAATATTAATGATAATTATAATCGACAAATGAACTTGACTTCATTATACACAAATCAT 1500
C V D F Y N Q G L N E A K A L S N S N W
TGCGTTGATTTTTATAATCAGGGTCTAAATGAAGCTAAAGCATTATCAAATTCAAATTGG 1560
D I F N D Y R R E M T I T V L D I V A L
GATATTTTTAATGATTATCGTAGAGAAATGACTATAACAGTTTTAGATATAGTTGCTCTA 1620
F S S Y D Y R R Y P I T T K V E L T R E
TTTTCTTCCTATGATTACCGTCGCTATCCTATAACTACAAAAGTAGAACTTACTAGAGAG 1680
I Y T P A I A S Q T W S N H N H L S P N
ATATATACACCTGCAATTGCTTCTCAGACTTGGAGTAATCATAATCATCTGTCACCGAAT 1740
I N F E F Y E N N L V R P P A F F T W L
ATAAATTTCGAATTTTACGAAAATAACCTTGTGCGGCCTCCTGCTTTTTTTACTTGGCTA 1800
D R T E M F S R Y L S T V V S E A W G G
GATAGAACCGAGATGTTTTCTAGGTACTTAAGTACTGTGGTTTCTGAAGCTTGGGGAGGG 1860
H I N H F H H T G E L P L S S R S G F I
CATATAAATCATTTTCATCACACTGGAGAACTACCTTTATCTTCACGTAGTGGTTTCATT 1920
G S D Q R R V A Y Y D F N F V G N D V F
GGAAGTGATCAGCGCCGAGTCGCATACTATGATTTCAATTTTGTTGGTAATGATGTCTTT 1980
R I Y S R V M S N P V G N Y F G V G Q G
CGTATATATTCAAGAGTAATGTCAAATCCAGTTGGGAACTACTTTGGCGTAGGACAGGGA 2040
D F Y L V N R D N C N T K T I T F T T K
GACTTTTATCTTGTGAATAGAGATAATTGTAATACTAAAACAATAACTTTTACTACTAAA 2100
A T N S N Q R S I L S E F P G E N S D P
GCAACGAATTCAAATCAAAGATCCATATTGTCTGAATTTCCGGGTGAAAATTCAGACCCA 2160
P T S K D Y S H R L S W I S G A F I G S
CCAACTTCTAAGGATTATAGCCATAGATTATCATGGATATCAGGTGCATTTATTGGTAGC 2220
D I A N V L S Y G W T H R S V D P N N T
GATATTGCAAATGTGCTTTCATATGGCTGGACTCATAGAAGTGTTGACCCTAACAATACT 2280
I Y P D K I T Q I P A V K L S S A S N C
ATTTATCCAGATAAGATTACTCAAATTCCGGCTGTTAAACTAAGTAGTGCTTCCAATTGC 2340
T V I P G P G S T G G H L V S F D R N G
ACTGTAATCCCAGGCCCTGGATCTACGGGAGGTCATTTAGTAAGTTTTGATAGGAACGGA 2400
S L D M Q F E F I T T Q T E Y R I R I R
AGTTTGGATATGCAATTTGAATTTATAACTACACAAACAGAGTATCGTATTCGTATACGC 2460
Y A S I A I N T L F F S F S G V N Q S I
TATGCCTCTATAGCAATAAATACACTATTCTTTTCTTTTAGTGGAGTAAATCAAAGTATA 2520
A L N S T G A S S L N N L R S E D F A Y
GCACTTAATTCTACAGGTGCTTCGTCACTAAATAATTTGCGAAGTGAGGATTTTGCGTAC 2580
L E F P Y G I F K P A I G N T L R I S N
TTGGAATTTCCGTATGGTATTTTTAAACCTGCTATAGGTAATACATTAAGGATTTCAAAC 2640
W S T V A P H L V I D K I E F I P I N S

TGGAGTACTGTTGCTCCACACTTGGTAATAGATAAAATTGAATTCATCCCAATTAATTCC 2700
T T A K Y E R M K E I E K A T T V V N S
ACTACTGCAAAATATGAGAGAATGAAAGAGATAGAAAAAGCCACAACAGTAGTGAATAGT 2760
L F I N *
TTATTTATAAATTAATACATGATATTAAAATACATGCCCTTGCTATTTAAAAATTGAACG 2820
RBS M F T N N A E N T L K I E T
GAAAAGGTTGTGGGAAATATGTTCACTAATAATGCGGAAAATACATTGAAAATAGAAACA 2880
T D Y E I D Q A A I S I E Y M S D E Q Y
ACAGATTATGAAATAGATCAAGCGGCTATTTCTATAGAATACATGTCGGACGAGCAATAT 2940
P Q E K M M L W E E I K H A K Q L S E S
CCACAAGAAAAAATGATGTTATGGGAAGAAATAAAGCATGCAAAACAACTTAGCGAATCT 3000
R N L L Q N G D F Q D S Y G Y G E N G W
CGTAATTTACTACAAAATGGAGATTTTCAAGATTCTTATGGGTACGGGGAAAATGGGTGG 3060
T N S N G I T I Q S N D P I F K G H Y L
ACAAACAGTAATGGTATTACCATTCAATCTAATGATCCTATTTTTAAAGGACATTATCTT 3120
Q M F G A R N I D G T L F P T Y I Y Q K
CAAATGTTTGGGGCAAGAAATATTGATGGAACGCTATTTCCAACCTATATCTATCAAAAA 3180
I D E F K L K P Y T R Y R V R G F V R S
ATAGATGAATTTAAATTAAAACCATATACACGTTATCGAGTAAGAGGATTTGTAAGAAGT 3240
S K D L K L V V T R Y G K E I D V I M D
AGTAAAGATTTAAAATTAGTGGTAACACGCTACGGTAAGGAGATTGATGTTATTATGGAT 3300
V P N D V A Y M Q P R H S C G D Y N R W
GTTCCAAATGATGTGGCATATATGCAACCACGTCATTCATGTGGAGATTATAATCGTTGG 3360
E S L S Q S V M N Q E Y P T P Y A A D A
GAATCATTGTCCCAGTCTGTGATGAACCAAGAATATCCTACACCATATGCAGCAGATGCC 3420
F D M Y S S Q F N R G K K H V T C H D C
TTCGATATGTATTCATCCCAGTTCAATCGAGGTAAGAAACATGTTACGTGTCACGATTGT 3480
H S F D F H I D I G E L D T N T N L G I
CATTCATTTGATTTTCATATTGACATAGGAGAATTAGATACAAATACAAACTTAGGTATT 3540
W V L F K I S N P D G Y A T L G N L E V
TGGGTCTTATTTAAAATTTCTAATCCAGATGGATACGCTACATTAGGCAATCTAGAAGTA 3600
I E E G P L T D E T L A H V K Q K E K K
ATTGAAGAAGGACCACTAACAGACGAAACATTAGCACATGTGAAACAAAAGGAAAAGAAA 3660
W N Q Q M E K K R C E T Q Q A Y N R A K
TGGAATCAACAGATGGAGAAAAAACGTTGTGAAACACAACAAGCCTATAATCGAGCAAAA 3720
Q A V D R L F T S T Q G E E L Q Y H I T
CAGGCAGTAGATAGATTATTCACAAGTACACAAGGAGAAGAATTACAATATCATATTACT 3780
L D H I K K S D Q L V Q S I P Y V H Q D
TTAGATCATATTAAGAAATCCGATCAGTTGGTACAGTCGATTCCCTATGTACATCAGGAT 3840
W L S D V P G M N A D L Y T D L N G R I

```
TGGTTATCAGATGTTCCAGGTATGAACGCTGATTTATATACAGATTTAAATGGACGTATA    3900
T  Q  A  R  Y  L  Y  D  A  R  N  I  I  T  N  G  D  F  T  Q
ACGCAAGCACGTTATTTGTATGATGCACGAAACATTATAACAAATGGTGATTTTACACAG    3960
G  P  T  G  W  S  A  S  G  H  E  A  F  K  K  I  D  G  D  S
GGACCAACAGGATGGAGCGCATCAGGACACGAGGCGTTCAAAAAAATAGATGGAGATTCT    4020
V  L  V  L  S  S  W  S  T  G  V  S  Q  N  L  H  V  Q  H  H
GTATTAGTTCTATCAAGCTGGAGTACCGGGGTATCTCAAAATCTGCATGTGCAACATCAT    4080
H  G  Y  V  L  R  V  I  A  K  K  E  G  L  G  K  G  Y  V  T
CATGGGTATGTATTACGTGTGATTGCGAAAAAAGAAGGGTTAGGAAAAGGATATGTAACG    4140
M  M  D  C  N  E  N  Q  E  T  L  K  F  T  S  C  E  E  G  Y
ATGATGGATTGTAATGAAAATCAGGAAACGCTTAAATTCACTTCTTGTGAAGAAGGATAT    4200
I  I  K  S  V  E  V  F  P  E  S  D  C  I  R  I  E  I  G  E
ATAATAAAATCAGTAGAGGTATTCCCAGAAAGCGATTGTATAAGAATAGAAATTGGAGAA    4260
T  E  G  T  F  Y  I  Q  S  I  E  L  L  C  M  K  G  Y  T  G
ACCGAGGGTACGTTTTATATACAAAGTATCGAGTTGCTTTGTATGAAAGGTTATACTGGA    4320
N  C  N  *                IR                    IR
AATTGTAACTAAAATACGCGTGCTATGTATGAACAAATATATAGTAGCAACTACAACCAT    4380
AATACTAGCGATAGGTATAATCAAAATTATACCAACAATTACGACCAGCATTCCAGCTGT    4440
ACGTGTAATCAAGAATATAACCGTTAAGATTCAAAATGAGGATCAGCATATTGACGAAAA    4500
AAATAAAAACCTACTCACAAAATCTATTGCGTATCATAACATAAGCTTTACAAATAGGAG    4560
ACATATTCTAGAACTGGTCTCCTTAATTCTAAAATAAGGAGGTCCTTTTCGTTTTCACAA    4620
TATCGATTAATGAAAATACTCCTTTACAGAACGATTTAGGCTGATTAGATTTGAATGTTG    4680
TTGAATTG                                                        4688
```

Fig. 27 Nucleotide sequence and deduced amino acid sequence of the *cry73Aa* and *cry40orf2* gene. The *cry73Aa* gene is 2,010 bp in length and codes for a polypeptide of 669 amino acids. The *cry40orf2* gene is 1,494 bp in length and codes for a polypeptide of 497 amino acids. The potential -35 and -10 boxes and a putative ribosome-binding site (RBS) are overlined. The stop codon is marked with asterisks. Five conserved sequence blocks (blocks 1 to 5) are shadowed. Terminal inverted repeats (IR) are indicated below the arrow.

Translation map – *cry20Bb1*

```
CTTGATTTCGAGGGAAACAAGCTAAATTATTATTTGAGTAAGAAACGAGATGCGAAGGCT    60
GCCAAGCGTTTTCTACAAAAACCCTTAGCTTCTTTTCATGTCACAAAAACTCGTGTCATA   120
ACTGCAGATGGTGATAAAGCCTATCCCGTTGTGATATAGAAATTAATAATGAAAAAAGCA   180
        -35                     -10
TACTGGTGGATAAATATAAATAAAACCTAATATATAAAAATAAATATACGAAGAGGGATT   240
CTAAAAAATCAATACTTTTACCAAAAATAATAGCTTTATTTGTAGAAAGATTATTTCAGG   300

AATGCATAGGCGCATTACGAATATTTACAAAGAAACCAATTATATATTTGTTAAAGGTAT   360
                    RBS          M  N  S  Y  K  N  N  H  T  M
TAAAAAATGTCTACATAAGGAGTGAAAAAAAATGAATTCTTATAAAAATAACCATACAAT   420
 V  N  S  P  E  N  S  S  N  T  V  N  R  Y  P  Y  A  C  N  P
GGTTAATTCCCCCGAAAATTCTAGCAATACTGTAAATAGGTATCCTTACGCTTGTAATCC   480
 N  I  E  T  Q  N  M  N  Y  K  D  W  M  A  G  Y  E  E  I  A
AAATATTGAAACGCAAAATATGAATTATAAAGATTGGATGGCTGGATATGAAGAAATTGC   540
 P  S  S  L  S  L  I  L  S  S  I  G  I  L  N  Q  V  I  A  L
TCCATCTTCATTATCTTTAATTTTATCTTCAATAGGCATTCTTAATCAAGTAATTGCCTT   600
 T  G  V  L  G  K  T  P  E  I  I  N  I  V  Q  E  M  V  S  L
AACTGGCGTATTAGGTAAGACACCAGAAATTATTAACATAGTACAAGAAATGGTGAGTTT   660
 I  R  G  N  T  G  N  D  L  L  V  H  V  E  Q  L  I  Q  Q  T
AATTAGAGGGAACACAGGCAATGATTTATTAGTACATGTAGAACAACTTATTCAACAAAC   720
 L  A  T  Q  Y  R  S  A  A  T  G  A  I  Y  G  I  S  R  A  Y
ATTGGCAACACAGTATAGGAGCGCTGCAACTGGAGCCATATATGGTATATCTAGAGCATA   780
 N  N  Y  L  E  F  F  R  Q  W  E  R  N  R  T  P  Q  N  G  Q
CAATAATTATTTGGAGTTCTTTAGGCAATGGGAACGTAATAGAACTCCTCAAAATGGGCA   840
 Q  V  E  S  A  F  T  T  V  N  T  L  C  I  N  A  L  A  P  Q
ACAAGTCGAGAGTGCTTTTACTACTGTTAATACTTTATGTATTAATGCTTTAGCCCCTCA   900
 A  S  L  S  R  R  G  F  E  T  L  L  L  P  N  Y  A  L  A  A
GGCGTCACTTTCACGCAGAGGATTCGAAACTCTTTTATTACCCAACTATGCTCTAGCGGC   960
 N  F  H  L  L  L  L  R  D  A  V  L  Y  R  T  Q  W  L  A  N
AAATTTCCATTTGTTATTATTAAGAGATGCTGTTCTTTATAGAACCCAGTGGTTAGCTAA  1020
 S  I  S  T  T  N  V  N  I  Q  I  L  T  R  A  I  N  E  Y  R
TTCTATTTCAACAACAAATGTAAATATCCAGATATTAACAAGAGCCATAAATGAATATCG  1080
 N  H  C  N  Y  W  Y  N  N  G  L  N  R  F  T  R  T  S  F  N
TAATCATTGTAATTATTGGTATAATAACGGATTAAATAGATTTACACGCACCTCTTTTAA  1140
 D  W  V  R  F  N  A  Y  R  R  D  M  T  L  S  V  L  D  F  V
TGATTGGGTTCGATTCAATGCTTATCGTAGAGATATGACGTTGTCGGTATTAGATTTTGT  1200
 T  V  F  P  T  Y  D  P  I  R  Y  P  R  P  T  N  V  E  L  T
TACAGTATTTCCAACGTATGATCCTATTCGATATCCAAGACCAACAAATGTTGAGTTGAC  1260
```

R I V Y T D P I S P P R G F P R T N P P

TAGAATTGTTTATACCGATCCAATAAGTCCACCTAGAGGATTTCCTAGAACCAATCCTCC 1320

S F N Q M E N L I I S G G P S F L N Q L

TTCCTTTAATCAGATGGAAAATTTAATTATTTCGGGTGGTCCTAGTTTCTTGAATCAATT 1380

R I Y T T F Y H D P H H V N R D Y W A G

GAGAATATATACAACTTTTTATCATGATCCTCATCATGTAAATAGAGACTATTGGGCCGG 1440

N R N Y L S N G I S R Q S G S T T P W P

GAATCGGAATTATTTAAGCAATGGGATTTCTCGACAGTCTGGATCTACCACACCTTGGCC 1500

T N I P M Q N I D I F R V N L T T H D I

AACTAATATACCTATGCAAAATATTGATATTTTCAGAGTGAATCTAACTACCCATGATAT 1560

D S I Q R S Y G G V H R S D F I G V N T

TGACTCTATACAACGAAGTTATGGCGGAGTTCATAGATCTGATTTCATTGGTGTAAATAC 1620

I N N Q R T T L F Y H Q N V D T S R F L

AATAAATAATCAAAGAACAACATTGTTCTATCACCAAAATGTGGATACTTCTCGTTTTCT 1680

T R N E T V F L P G D S G L E P N E Q N

AACAAGGAATGAAACAGTATTTTTACCAGGGGATTCAGGATTAGAACCAAATGAACAGAA 1740

Y T H R L F Q V M T T Y R I N P N A R R

TTATACTCACAGGTTATTTCAAGTCATGACCACATATCGTATTAACCCGAATGCTCGTAG 1800

A A F L H A W T H R S L R R R N G F R T

GGCAGCTTTTTTACATGCATGGACGCATAGAAGTTTAAGACGTAGAAATGGATTTAGGAC 1860

D Q I M Q I P A V K T I S T G D D R A V

GGATCAGATTATGCAAATACCAGCTGTTAAGACCATAAGCACTGGTGATGATCGTGCTGT 1920

V L N Y G E N I M K L D N L T S G L S Y

AGTGTTAAATTATGGAGAAAACATCATGAAATTAGATAATTTAACTTCAGGTTTATCCTA 1980

K V T A T D S A A S N T R F I V R V R Y

TAAAGTAACGGCGACAGATTCAGCAGCGTCCAATACACGTTTTATTGTGCGTGTTCGTTA 2040

A S M D N N K L N L V L N G A Q I A S L

TGCTAGTATGGACAATAATAAATTGAATCTTGTTTTAAATGGCGCTCAGATAGCATCACT 2100

N V E R T V Q N G G E S L T A L Q C E D

AAATGTGGAACGCACAGTGCAAAATGGCGGCGAATCATTAACAGCTCTTCAATGTGAAGA 2160

F K Y A T F A G D F Q M G S Q S I F G I

TTTTAAATATGCTACATTTGCAGGTGATTTCCAAATGGGTTCTCAATCTATATTCGGTAT 2220

F K D I S N A D F I L D K I E L I P S H

TTTTAAAGACATATCTAATGCAGACTTTATTTTAGATAAAATTGAATTAATCCCATCCCA 2280

F M S S L E Q T Q D D Y S Y N Q N T I Y

TTTCATGTCATCATTAGAGCAAACACAAGATGATTACAGCTATAATCAAAATACTATTTA 2340

T C N Q G Y G T Y D H N S S N M Y D H Q

TACATGTAATCAAGGATATGGTACTTACGACCATAATTCTAGTAATATGTATGATCATCA 2400

N Y K N Y T Q D M D T T Y Q P D Y D N Y

AAACTACAAAAATTATACTCAAGACATGGATACTACATACCAACCAGACTATGATAATTA 2460

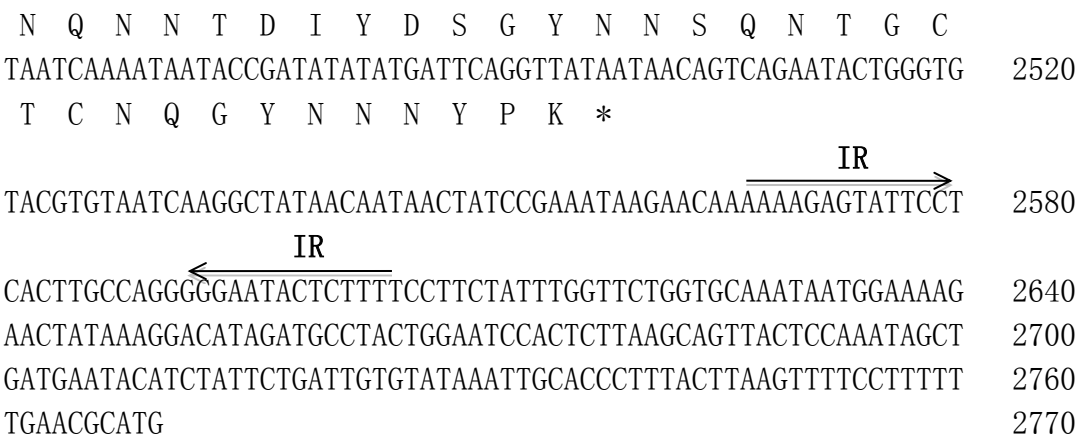


Fig. 28. Nucleotide sequence and deduced amino acid sequence of the *cry20Bb1* gene. The *cry20Bb1* gene is 2,178 bp in length and codes for a polypeptide of 722 amino acids. The potential -35 and -10 boxes and a putative ribosome-binding site (RBS) are marked. The stop codon is marked with asterisks. Five conserved sequence blocks (blocks 1 to 5) are shadowed. Terminal inverted repeats (IR) are indicated below the arrow. The N-termini of 50 kDa (LLVHV) and 30 kDa (NVNIQ) degradation products are double underlined.

Translation map - *cry27Ab1*

```
CCAGCTACAAAAATGATTGTTGGAATAGAGGCCATGCATATAGTCAAAAAAGGTCAACTA     60
AAATTAAGGGTACAACCTGACAAAAATCAGAATAGATGTATTCATTAGTTATTTGGATTA    120
ACTGCTTAAGAGTGGATCCCGTTAGGAATCTATGCCTTTTGTAGCCTTTGTCCATTATTT    180
GCAACAGAACCCCTTTATCCTAACCACCATACTCACATCCAAAAATAATGGGTTTATTTG    240
            -35                 -10
TAGAAAGATTGTTACAGGAATACATAGGTACATTACGAATTTTTAAGAAAGACACCTACT    300
                                RBS       M  N  P  Y  Q  D  K
ATATTTATAGAGGTGGCATAAAGACTAAGGGAGGAACCAAAATGAATCCTTATCAGGATA    360
  N  E  Y  K  I  L  D  A  K  R  N  T  C  H  M  S  N  C  Y  P
AGAATGAATATAAAATCTTAGATGCTAAACGAAATACTTGTCACATGTCTAATTGTTATC    420
  K  Y  P  L  A  N  D  P  Q  M  Y  L  R  N  T  H  Y  K  D  W
CCAAATACCCATTGGCAAATGATCCTCAAATGTATTTGCGCAATACTCATTATAAGGATT    480
  L  T  M  C  N  N  T  K  L  A  G  W  I  P  P  G  S  F  E  F
GGCTAACTATGTGTAATAATACCAAGCTTGCAGGTTGGATACCGCCAGGGAGCTTTGAGT    540
  T  W  L  N  A  T  V  A  A  L  T  I  I  S  V  T  T  A  L  F
TTACCTGGCTAAATGCAACTGTCGCTGCACTTACTATCATTAGTGTAACTACAGCTTTAT    600
  I  A  P  P  L  L  V  G  G  V  I  A  A  G  A  A  I  L  A  G
TTATAGCTCCACCTCTTCTGGTAGGAGGTGTTATAGCTGCAGGAGCTGCTATTTTAGCAG    660
  T  L  P  L  L  W  P  A  D  S  K  P  E  D  N  T  F  N  E  I
GTACATTACCTCTTTTATGGCCTGCGGATTCTAAACCTGAAGATAATACATTTAATGAAA    720
  M  N  A  T  E  V  L  L  N  T  K  I  S  D  F  V  R  Q  T  A
TTATGAATGCAACAGAAGTTCTACTTAATACAAAAATATCTGACTTTGTTAGACAAACAG    780
  D  T  K  I  T  S  L  Q  S  L  M  F  Y  Y  N  N  A  L  D  N
CAGATACCAAAATTACTAGTTTACAAAGTTTAATGTTTTATTATAACAACGCTTTAGATA    840
  W  K  K  N  P  N  D  S  A  A  I  N  T  V  S  T  R  F  Q  I
ATTGGAAGAAAAACCCGAATGATTCAGCCGCGATAAATACGGTAAGTACTAGGTTTCAAA    900
  V  N  A  F  F  V  E  A  M  T  A  L  S  M  P  G  Y  E  L  A
TTGTGAATGCTTTTTTTGTCGAAGCTATGACAGCCCTTTCTATGCCAGGATATGAATTAG    960
  Q  L  G  A  Y  A  Q  A  A  N  L  H  L  L  L  L  R  D  G  I
CACAATTAGGTGCATATGCACAAGCAGCTAATCTGCATTTATTACTTTTACGAGATGGAA   1020
  L  Y  A  D  K  W  N  L  A  K  E  A  T  Y  K  Q  G  D  L  H
TTTTATATGCAGATAAATGGAATTTAGCCAAAGAAGCAACCTATAAACAAGGAGACTTAC   1080
  Y  Q  E  F  L  N  Y  R  N  Q  Y  I  N  H  C  S  T  W  Y  T
ATTATCAAGAATTTCTAAATTATAGAAATCAATATATTAACCATTGTTCAACTTGGTATA   1140
  E  G  Q  I  E  A  N  N  K  G  N  G  L  V  Y  Q  R  T  M  T
CTGAAGGACAAATAGAGGCAAATAATAAAGGTAATGGACTTGTATATCAAAGAACTATGA   1200
  I  L  V  L  D  L  I  A  M  F  S  T  Y  D  P  R  L  Y  T  M
CAATTTTAGTACTAGATTTAATTGCAATGTTTTCAACATACGATCCACGCTTATATACGA   1260
```

P T K T E I L T R T L Y T D G V N R N Q
TGCCAACTAAAACCGAAATTTTAACAAGAACACTTTATACAGATGGAGTAAATAGAAATC 1320
T R S I H N P G L F R R L E Q M E L H T
AAACTAGATCTATACACAATCCAGGTTTATTCCGGCGGTTAGAACAAATGGAATTACACA 1380
Y E Y Q G A Q F L S G H Q N K F R S M N
CTTATGAATATCAGGGTGCACAGTTTTTAAGTGGGCACCAAAATAAATTTAGAAGCATGA 1440
Y N H P L I D G P V Q G Y S S S N I N K
ATTATAATCATCCTCTAATTGACGGTCCCGTACAAGGGTATAGTTCATCGAATATAAATA 1500
I T T I N L G D Y D K I Y S I K T E S R
AAATAACAACTATTAATCTAGGTGATTATGATAAAATTTATAGTATTAAAACAGAAAGCA 1560
D R I V Q G S I T F D K I N F Y G A F N
GAGATCGTATAGTTCAGGGCTCAATTACATTTGATAAAATTAATTTCTATGGGGCATTTA 1620
K S W L F S V Y N Q N G P I I K H S N I
ATAAAAGTTGGCTATTTTCTGTATACAATCAGAACGGTCCAATTATAAAACACAGCAATA 1680
P G V E A P S A T L D Y R N Y T H Y L S
TACCAGGTGTTGAGGCCCCTTCAGCAACGTTAGATTACAGAAACTATACTCATTATTTAT 1740
N C I F Q S N R N G V S E P D Y N T Q S
CAAATTGTATCTTCCAATCAAACCGAAATGGAGTATCTGAACCAGATTACAACACCCAAT 1800
Y I F G W N H N T I D P T G N Y V T D A
CATATATATTTGGCTGGAATCACAATACTATTGATCCCACAGGGAATTATGTAACAGATG 1860
S F V D N G L P E G R Y V P Q I S Q V P
CAAGTTTCGTAGATAACGGCTTACCTGAAGGACGATATGTACCCCAAATTTCACAAGTGC 1920
A V K A S D I Y N P G R V V N A T V E V
CCGCTGTAAAAGCTAGTGATATATACAACCCAGGTCGTGTAGTTAATGCAACAGTTGAAG 1980
G P Y F T G G D V I V S K A Q L D G S G
TTGGACCATATTTTACAGGTGGCGATGTTATTGTATCGAAAGCTCAATTAGATGGATCAG 2040
L A R T V I T F P I I P K R Y Q A S G F
GTCTAGCCAGAACAGTTATAACATTCCCTATTATACCGAAAAGGTATCAAGCAAGCGGAT 2100
R V R M Y Y A A N H I G Q L S Y R A K D
TTCGTGTACGTATGTACTATGCTGCCAATCATATTGGTCAATTGAGTTACCGTGCGAAAG 2160
L N V T G Y A N F T K T F D G W E Y F R
ATCTAAATGTAACCGGTTATGCAAATTTTACAAAAACATTTGATGGCTGGGAATATTTTA 2220
A R H E H F K Y I E F D T T F S L R N S
GAGCGCGACACGAACATTTTAAATATATAGAATTTGATACGACATTTAGCTTACGAAATT 2280
G Q L E E H E L Q I Y Y P N T S R V S G
CAGGTCAATTAGAGGAACATGAATTACAAATTTATTATCCCAATACTTCAAGAGTATCTG 2340
D Q L L I I D K I E F I P V G I R L N Q
GCGATCAATTATTAATTATAGACAAAATTGAATTTATACCGGTGGGAATTCGACTAAATC 2400
P S E G Y N T Y D Q N T N S Y N Q N Y N
AACCATCAGAAGGATATAATACGTACGATCAGAATACTAATAGCTATAATCAAAACTATA 2460

```
  N  Y  N  Q  N  M  D  T  T  Y  Q  P  N  Y  D  N  Y  N  Q  N
ACAATTACAACCAAAACATGGATACTACATATCAACCAAACTATGACAATTATAATCAGA    2520
  S  S  G  M  Y  D  N  P  Y  N  Q  N  P  K  D  S  Y  N  Q  N
ATTCCAGTGGTATGTATGACAATCCCTATAACCAAAATCCCAAGGATAGTTATAATCAAA    2580
  Y  T  D  T  Y  D  S  G  Y  N  N  S  Q  N  V  G  S  N  Y  N
ATTATACCGATACGTATGATTCAGGATATAATAACAGCCAGAACGTTGGCTCTAATTATA    2640
  Q  E  Y  N  T  Y  N  Q  D  T  E  N  M  Y  N  Q  S  Y  N  N
ATCAAGAATACAATACTTATAACCAAGATACGGAAAATATGTATAACCAATCGTATAACA    2700
  Y  N  S  D  N  N  N  Y  N  Q  N  S  D  C  V  C  S  P  G  Y
ATTATAATTCTGACAACAACAATTACAACCAGAATTCTGATTGCGTGTGTAGTCCAGGAT    2760
  N  G  N  Y  E  C  R  C  N  Q  R  A  N  G  N  Y  P  K  *
ATAATGGTAACTATGAATGTAGATGTAACCAAAGGGCTAACGGTAATTATCCTAAATAAG    2820
                                IR                IR
AACAGTAAAAACATACTAAGATTAGTAGTGCGGACCAAGGCATTGGTTCGCACTACTATT    2880
TTTTATATAGTAGGGGTCACCATACATGCCCATCAACTTAAGGATGGATACAAAACAAAA    2940
TTTTGTTCATTTGAACTGAAATCTCTCCAAATGGGGGTAAACCCAGCATCTTCGAATGAA    3000
CCATTAATTTCGTAATTTTCTAATAATTAATAGGTATTTTCATGTTTTTTGGGCTTGAAA    3060
TCCAAGTG                                                        3068
```

Fig. 29. Nucleotide sequence and deduced amino acid sequence of the *cry27Ab1* gene. The *cry27Ab1* gene is 2,478 bp in length and codes for a polypeptide of 825 amino acids. The potential -35 and -10 boxes and a putative ribosome-binding site (RBS) are marked. The stop codon is marked with asterisks. Five conserved sequence blocks (blocks 1 to 5) are shadowed. Terminal inverted repeats (IR) are indicated below the arrow.

Translation map – *cry4Aa*

```
                                                  -35
GAATCGCGCATATTTCAGGTGAATTTTATTTCGAATATGAATAATACATATTTTCCTTAC     60
-10
TTATACTCCTATCCCTACCCCAAAAAAGAATCTATGCTCAGATTCTTTTTTATCTTTTGT     120
CTACAGGAATAGTCAATCTTTTATCCAAATAAATATAAGATTGGAAATTAAATTGAAAGT     180
RBS          M  N  H  Y  N  N  E  N  Y  E  I  I  D  S  N  T
GGAGGAACATATATGAATCACTATAACAATGAAAACTATGAAATTATCGATTCCAATACT     240
 S  P  Y  P  S  N  R  N  N  P  Y  S  R  Y  P  Y  A  N  N  P
TCCCCGTATCCTTCTAACAGAAATAATCCATATTCTAGATATCCTTATGCAAATAATCCT     300
 N  Q  S  L  Q  N  K  N  Y  K  D  W  M  S  I  P  Q  P  Y  A
AATCAATCATTGCAAAATAAAAATTATAAAGATTGGATGAGTATACCTCAACCTTATGCA     360
 D  C  D  N  N  S  F  D  W  L  A  A  V  S  A  G  V  I  V  I
GATTGTGATAACAATTCATTTGATTGGCTTGCCGCTGTTAGTGCAGGTGTCATTGTAATA     420
 G  T  M  L  A  A  F  A  A  P  I  A  A  P  V  L  A  G  S  I
GGTACTATGTTAGCTGCTTTTGCTGCTCCTATAGCAGCGCCAGTTCTAGCTGGATCTATT     480
 I  I  S  I  G  T  L  L  P  I  L  W  P  L  G  Q  S  D  N  N
ATTATATCAATCGGTACATTACTTCCTATTCTTTGGCCACTTGGTCAATCAGATAATAAC     540
 A  V  W  Q  K  F  L  D  Q  G  N  N  L  T  C  Q  Q  L  T  P
GCAGTATGGCAAAAATTCCTTGATCAAGGAAATAACCTTACATGTCAACAACTAACTCCA     600
 G  I  K  V  A  V  D  A  A  L  N  N  L  R  V  Q  A  H  Y  F
GGAATTAAAGTAGCAGTAGATGCAGCTTTAAATAATTTAAGAGTTCAAGCCCATTATTTC     660
 N  D  A  V  T  Y  W  E  K  S  I  G  T  S  N  E  I  D  A  R
AACGATGCCGTTACCTATTGGGAAAAAAGTATAGGCACTTCAAATGAAATAGATGCTAGA     720
 N  N  A  R  D  I  Y  I  N  A  V  Q  I  I  E  G  L  M  P  I
AATAATGCAAGAGATATTTATATAAATGCTGTACAAATAATTGAAGGACTTATGCCTATA     780
 F  K  S  S  G  Y  E  V  L  L  L  S  T  Y  A  Q  A  A  L  L
TTTAAATCATCAGGTTATGAAGTATTATTATTATCTACTTATGCACAAGCCGCATTATTA     840
Q  V  T  L  L  H  Q  G  I  Q  Y  A  S  K  W  N  L  A  R  D
CAAGTTACTTTACTACATCAAGGCATTCAATATGCTTCCAAATGGAATTTAGCTCGAGAT     900
T  G  D  F  Y  R  Q  R  L  Y  E  A  I  D  R  H  I  D  Y  C
ACTGGAGATTTTTATCGTCAAAGACTTTACGAAGCAATAGATAGACACATTGATTATTGT     960
E  T  W  Y  Q  T  G  L  D  E  L  K  K  N  E  N  L  T  F  A
GAAACATGGTATCAAACAGGTCTAGACGAACTCAAGAAAAACGAAAATTTAACATTTGCT    1020
A  Y  I  N  Y  R  R  E  Y  T  I  N  V  L  D  V  I  S  L  I
GCCTATATAAATTATCGTAGAGAATATACTATCAATGTATTAGATGTTATTTCCCTAATT    1080
P  A  L  D  L  R  I  Y  P  D  T  K  P  I  N  I  E  F  T  R
CCAGCATTAGATTTACGTATTTATCCAGACACTAAACCAATTAACATAGAATTCACGCGA    1140
N  I  F  T  A  I  P  T  S  N  Q  S  R  I  S  A  F  I  G  R
```

AATATATTTACAGCTATACCAACTTCAAACCAATCAAGAATAAGTGCATTTATAGGACGT 1200
E N I E K L E K E L W P S T E L F T Q L
GAAAACATTGAAAAATTAGAAAAAGAACTATGGCCTTCTACAGAATTATTTACACAGTTA 1260
R Q I A F Y Q D Y N Y I E S G N Y L A Q
AGACAAATAGCTTTTTATCAAGATTATAATTATATCGAAAGCGGAAACTACTTAGCTCAG 1320
I T N L I V H A N D S N L I T K T Y G N
ATAACTAATCTAATCGTTCACGCTAATGATTCTAACCTCATTACAAAAACATATGGAAAC 1380
A P S Q T A S P T I V L S P Q Q S I Y N
GCACCATCACAAACAGCATCTCCAACTATAGTTCTCTCCCCACAACAAAGCATTTATAAT 1440
C T I D N Y S I G D T P R T G I K Y M E
TGTACAATAGATAATTATTCAATTGGAGATACACCAAGAACAGGAATAAAATATATGGAG 1500
L K V A T Q R I T S N S V K F G S E T S
CTAAAAGTAGCTACTCAAAGAATTACCAGTAATTCTGTAAAATTCGGTTCGGAAACTAGT 1560
G S Y Q R N I L P F P T D I Q V T S L Q
GGGTCTTACCAAAGGAATATATTACCGTTTCCAACAGATATACAAGTAACAAGTTTACAA 1620
N Y Q Y K L S R I T M S Q N K Y P M T S
AACTATCAATATAAATTATCTCGTATCACAATGTCTCAAAATAAATATCCCATGACTTCG 1680
V G E T T T Y L Y G F I W T H A Q S N P
GTCGGCGAGACTACTACTTATCTATATGGATTTATTTGGACACACGCTCAATCTAATCCT 1740
T N T I T S K N K N N Q K T I T Q I S A
ACAAATACCATTACTTCTAAAAATAAAAATAATCAAAAAACAATTACACAAATCTCTGCA 1800
V K A Y E L S N P N S H I F P N T I T V
GTTAAAGCATATGAACTTTCCAATCCTAATTCTCATATATTTCCCAATACAATTACAGTT 1860
I E G P G H T G G K L V K S T Y I L D Q
ATAGAAGGACCTGGTCATACAGGTGGAAAACTAGTAAAATCCACCTATATTCTCGACCAA 1920
L V I K C T F T D S S Q Y R L R I R Y A
TTAGTAATTAAGTGTACATTTACTGATAGTAGCCAATATCGATTGCGTATTAGATACGCT 1980
T D I I N N G I L K V T I Q S S N N L N
ACAGATATAATAAATAATGGAATCTTAAAAGTAACTATACAATCTTCAAATAACTTAAAT 2040
I I K R Y E F S L R K G N L N T S T N I
ATTATTAAAAGATATGAATTTAGTCTAAGAAAAGGAAACTTAAATACATCTACAAATATC 2100
P L Y K D F L T T E A L N P F S V T A N
CCCTTATATAAAGATTTCTTAACAACGGAAGCTCTCAATCCATTTTCAGTTACAGCAAAT 2160
E K V N I I I E N G S T N S G T I L I D
GAAAAGGTAAATATAATAATAGAAAATGGATCAACAAATTCAGGAACTATTCTTATTGAT 2220
K L E F V P Q * IR IR
AAACTTGAATTTGTCCCACAATAAACAACAAAAATGAAAGTTCTTTTTCTTTTTTCGGTT 2280
TACAAAATATAGATATTACTTTTAATATCTATATTTATTTGAATTTACATTTATAGAGGG 2340
AAATGTTTTCGTCATACCTTGATCTACAAACACATTCCATGTTGCCCCAATAACATTTGT 2400
ATCATTTAAACGTTTTAATTTAAATTCATATTTCATTAAATATTTTACATATCTAGTTGG 2460

CTCCGCACC 2469

Fig. 30. Nucleotide sequence and deduced amino acid sequence of the *cry4Aa* gene. The *cry4Aa* gene is 2,052 bp in length and codes for a polypeptide of 684 amino acids. The potential -35 and -10 boxes and a putative ribosome-binding site (RBS) are overlined. The stop codon is marked with asterisks. Five conserved sequence blocks (blocks 1 to 5) are shadowed. Terminal inverted repeats (IR) are indicated below the arrow.

Translation map – *cry56Ba1+cry39orf2*

```
CCATTGGTTATTGAACATAGAATGATGAACAGAGTTTTATTACATATAAGACAATAAAAA    60
ACGCTTCTTTTAAATCAAAAAGAAGCGTGTTCAGAAAAGGGAAATTTTCTGTACACAGGT   120
                                                      -35
AAAATTCCCATAATTCGGTTTTCCCTAGAAAAAATAAAATGAATCCAATACATGTTCGGA   180
                -10
TTTATTTTCTACAATTTCTTTCAATGACGAATAATTAAATATCTACAACAATCTATTTAG   240
TATCCCTCTCTTTTTTCGAGGGATAGGAAGAAACATCCAAGTATGAATTTTGTTCATATA   300
AAAGTGAATAACTCTTTCCGCACCCTTAAAAACAACAAAGAAAAAATCGTTCTACAGAAA   360
TCTGAAGCTTTTAAAAAAATACATGCAATACATAAAGAGAAGATTTGAAAATTAAATACC   420
TGACCAAATATAATGGGTTTATTTGTAGAAACATCGTTACAGGAATACATTGGGGTACTT   480
CGAATATATAGAAAGACACCTAGCATATATTTATTAGGTGTTTTAAAAATAAGGACTACA   540
   RBS          M  N  S  Y  Q  N  K  N  E  Y  E  I  L  D  T
TAAGGAGTGAAAAATATGAATTCATATCAAAATAAAAATGAATATGAAATATTGGATACT   600
S  P  N  S  S  T  M  S  T  R  Y  P  R  Y  P  L  A  K  N  P
TCACCAAACAGCTCTACTATGTCTACTCGTTATCCTAGATACCCACTAGCAAAAAATCCA   660
Q  I  S  M  Q  N  T  N  Y  K  D  W  I  N  M  C  T  N  N  T
CAAATATCCATGCAAAATACGAATTATAAAGACTGGATAAATATGTGTACAAATAATACC   720
L  I  P  I  E  P  L  D  L  T  W  Q  N  A  L  V  S  V  F  G
CTTATTCCTATAGAACCTTTAGACCTTACCTGGCAAAATGCTCTTGTTTCAGTCTTCGGT   780
I  A  S  A  V  A  A  L  L  A  A  P  I  T  G  G  T  S  I  A
ATCGCTTCAGCTGTTGCAGCATTGTTAGCAGCTCCAATTACTGGCGGAACATCTATAGCA   840
A  G  A  A  I  I  A  N  I  L  P  L  T  F  P  A  N  A  E  S
GCTGGAGCGGCTATAATAGCTAATATATTACCATTAACCTTTCCCGCTAATGCTGAGAGT   900
V  P  N  K  L  M  D  A  T  Q  E  L  L  G  P  L  E  E  Y  T
GTTCCGAATAAGCTTATGGATGCCACACAAGAATTACTTGGCCCTCTAGAAGAATACACT   960
R  N  R  A  N  S  E  L  L  S  L  S  S  Q  L  E  A  F  K  G
AGAAATAGAGCAAATTCGGAGCTACTCAGTTTGAGTTCACAGTTAGAAGCTTTTAAAGGT  1020
L  F  D  Y  W  L  A  D  R  Q  N  P  N  A  T  N  S  V  S  A
CTATTTGATTATTGGCTCGCTGACCGCCAAAATCCAAATGCAACTAATTCAGTTAGTGCT  1080
R  F  T  A  I  H  N  N  F  I  G  A  M  A  L  F  K  I  P  G
CGTTTTACTGCAATTCATAATAATTTTATAGGGGCAATGGCTCTTTTTAAAATACCGGGT  1140
Y  E  A  L  L  L  P  V  Y  A  Q  A  A  R  L  H  L  L  H  L
TATGAAGCCTTACTGTTACCGGTATATGCTCAGGCTGCACGTTTACATTTGCTTCATTTA  1200
R  D  G  I  T  Y  A  D  Q  W  Q  L  A  D  P  T  N  A  A  Y
AGAGACGGTATCACGTACGCTGATCAATGGCAGTTAGCTGATCCAACTAATGCAGCTTAT  1260
A  G  D  Y  H  Y  S  E  F  K  K  Y  S  A  Q  Y  A  D  H  C
GCGGGAGATTACCACTATAGTGAATTTAAGAAATATTCTGCACAATATGCAGATCATTGT  1320
E  L  V  V  N  N  Q  L  N  K  I  K  N  T  N  G  K  T  W  K
```

GAATTAGTAGTTAATAATCAACTAAATAAGATAAAAAATACAAACGGTAAAACATGGAAA 1380
D Y N E Y R R K M I L S V F D I V A E F
GACTACAACGAATATCGTCGAAAGATGATATTATCTGTTTTCGATATTGTTGCTGAATTT 1440
S T F D P I L Y K G A I N R E I L T R K
TCAACCTTTGATCCAATTTTATATAAAGGAGCGATAAATAGAGAGATTTTAACACGTAAA 1500
I Y T D P V N F T P G F S I A D D E N R
ATATATACAGACCCAGTTAATTTCACACCTGGTTTTTCAATTGCTGATGATGAAAATAGA 1560
Y T V R P S N V K Q L V A S T L F T N V
TATACAGTTAGACCGTCAAATGTTAAACAATTAGTCGCCTCCACACTATTTACTAACGTG 1620
A S A Q Y A G F I G N R N R Y L S L L G
GCATCTGCTCAGTATGCTGGATTTATTGGAAATAGAAATCGTTATTTAAGTTTATTAGGT 1680
G E P L E G P V I G K S V S E N V V A G
GGAGAGCCACTTGAAGGACCTGTAATCGGAAAATCAGTATCCGAAAATGTTGTAGCAGGT 1740
V P T N E S I Y E V G V N G Y P N D Y P
GTACCAACAAATGAATCGATTTATGAAGTTGGTGTAAATGGTTACCCGAATGATTATCCA 1800
R N I G L R W G S L T R F Q N Y Y A G S
CGTAATATAGGTTTGAGATGGGGTTCATTAACTAGATTTCAAAATTATTATGCTGGAAGC 1860
Q Y N L G G L T T V S V P P K N N A P I
CAGTATAATTTAGGGGGGTTAACTACGGTCTCTGTGCCACCTAAAAATAATGCCCCAATA 1920
N N T N F T H R L S D I I L P G N S G S
AATAATACTAATTTTACTCATCGATTATCAGATATAATTCTTCCTGGAAATAGTGGCTCA 1980
S F A W T H V E V N P T E N Y L S T D Q
TCTTTTGCATGGACTCATGTTGAGGTCAATCCTACAGAAAACTATTTATCAACAGATCAA 2040
I N L I S A T K T S T Y N S M W K G P G
ATTAATTTAATATCTGCTACAAAAACTTCAACATATAACAGTATGTGGAAGGGACCTGGA 2100
F I G G D L T S S D I A F G E Y L F Y N
TTTATAGGAGGAGATTTAACAAGCAGCGACATAGCATTTGGAGAATACTTATTTTATAAT 2160
F K Y K S P G S S A R F K I R L R Y G S
TTTAAGTACAAATCCCCTGGTAGCTCAGCTAGGTTTAAAATTCGTTTACGTTATGGATCT 2220
W G S Y G S V Y Y I L G N T T S P K T L
TGGGGTAGTTATGGGTCGGTATACTATATATTAGGGAATACCACTTCACCAAAGACTCTT 2280
F E N T R L D L N N Y K Y D Q F K V V E
TTTGAAAATACTAGATTAGATCTTAATAATTATAAGTATGATCAATTTAAAGTAGTAGAG 2340
L W G T A E N I T D N N L I I K V A F A
CTTTGGGGAACTGCAGAAAATATTACAGACAACAATTTAATTATTAAAGTAGCCTTTGCT 2400
N T G G S T G F Y L D R L E L I P M T G
AACACAGGAGGTAGTACTGGGTTTTATCTAGATAGATTGGAATTAATCCCTATGACAGGG 2460
M P T E Y T E P Q K L E T A Q K A V N D
ATGCCAACAGAATACACTGAACCGCAAAAATTGGAAACAGCACAGAAAGCAGTAAACGAT 2520
L F T N *

TTATTTACCAATTAAAAAAAAGTATGTAATGAAGTAGGTAGTAATCCTGTTCAAAAAATA 2580

RBS M Y T N A M K N T L K I E

CGCAGAAAAGGTAGTGAATCCTATGTATACCAATGCTATGAAAAATACATTAAAAATAGA 2640

T T D Y E I D Q A A I S I E C M S H E K

AACGACGGATTATGAAATAGATCAAGCGGCCATTTCTATAGAATGTATGTCACATGAAAA 2700

Y P Q E K M I L W D E V K Q A K Q L S Q

ATATCCGCAAGAAAAAATGATATTATGGGATGAAGTAAAACAAGCAAAACAACTCAGTCA 2760

S R N L L Y N G D F E D A S N G W K T S

ATCTCGTAATTTACTCTACAATGGGGATTTTGAAGATGCATCAAACGGATGGAAAACAAG 2820

Y T I E I R K N S P I F K G Q Y L H M F

TTATACGATTGAAATTCGAAAGAATAGTCCCATTTTTAAAGGGCAGTACCTTCATATGTT 2880

G A R D V L G E V F P T Y V Y Q K I D E

TGGTGCAAGAGATGTTTTAGGTGAAGTGTTTCCAACATATGTGTATCAAAAAATTGATGA 2940

S K L K P Y T R Y R V R G F V G S S K D

GTCTAAATTAAAACCATATACACGTTATCGAGTAAGAGGATTTGTGGGAAGTAGTAAAGA 3000

L K L A V T R Y G K E I D A I M D V P N

TCTAAAACTAGCGGTAACACGTTACGGGAAAGAAATTGATGCCATTATGGATGTTCCAAA 3060

D L A Y M Q P N P S C G D Y R C D S P S

TGATTTGGCCTATATGCAGCCTAACCCTTCATGTGGAGATTATCGCTGTGACTCACCATC 3120

Q S M M S H G Y P T P V T D G S A S N M

CCAGTCTATGATGAGTCACGGATATCCTACACCAGTAACAGATGGATCTGCTTCTAATAT 3180

Y A C P S D R V K K H V K C H D R H L F

GTATGCATGCCCGTCAGACCGAGTTAAAAAACATGTGAAGTGTCACGATCGCCATCTATT 3240

D F H I D T G E L D T N T N L G I L V L

TGATTTTCATATTGACACAGGAGAGTTAGATACAAATACAAACTTAGGTATCTTGGTCTT 3300

F K I S H P N G Y A T L G N L E V I E E

ATTTAAGATTTCCCATCCAAATGGATACGCTACATTAGGGAATCTAGAAGTGATTGAAGA 3360

G P L T D E A L E H V R Q K E K K W N R

AGGGCCACTAACAGACGAAGCATTGGAACATGTGAGACAGAAAGAAAAGAAATGGAATCG 3420

H I E K A R M E T Q Q A Y D P A K Q A V

ACACATAGAGAAAGCGCGAATGGAAACACAACAAGCTTATGATCCAGCAAAACAGGCAGT 3480

D A L F T S A Q E L H Y H T T L N H I K

AGATGCATTATTTACAAGTGCACAAGAGTTACACTATCATACTACTTTAAATCATATTAA 3540

N A D Q L V Q S I P Y V N H A G L P D A

GAATGCCGATCAGTTGGTACAGTCGATTCCCTATGTAAACCATGCTGGGTTACCGGATGC 3600

P G M N Y D L Y Q G L N A R I M Q A Y N

TCCAGGTATGAACTATGATTTATATCAAGGGTTAAACGCGCGTATCATGCAGGCGTACAA 3660

L Y D A R N V I T N G D F T Q G L Q G W

TTTATATGATGCACGAAATGTCATCACAAATGGTGACTTTACACAAGGATTACAGGGATG 3720

H A T G N A A V Q Q M D G A S V L V L S

GCACGCAACAGGAAATGCCGCGGTACAACAAATGGATGGCGCTTCTGTATTAGTTCTATC 3780
N W S A G V S Q N L H A Q D H H G Y V L
AAACTGGAGTGCCGGGGTATCTCAAAATCTGCATGCCCAAGATCATCATGGATATGTGTT 3840
R V I A K K E G T G K G Y V T M M D C N
ACGTGTGATTGCCAAAAAAGAAGGGACCGGAAAAGGGTATGTAACGATGATGGATTGTAA 3900
G K Q E T L T F T S C E E G Y M T K T V
TGGAAAGCAGGAAACACTTACGTTCACTTCTTGTGAAGAAGGATATATGACAAAAACAGT 3960
E V F P E S D R V R I E I G E T E G T F
AGAGGTATTCCCAGAAAGTGATCGTGTACGGATTGAAATAGGAGAAACCGAAGGTACATT 4020
Y I D S I E L L C M K G Y P S N Y N Q N
TTATATAGATAGCATAGAGTTACTTTGTATGAAAGGGTATCCTAGCAATTACAACCAAAA 4080
T D N M Y E Q S Y N G N Y N Q N T S D V
TACAGATAATATGTATGAGCAAAGTTATAATGGAAATTATAATCAGAATACTAGCGATGT 4140
Y H Q G Y T N N Y N K D S S S M Y N Q N
GTATCACCAAGGATATACAAACAACTATAACAAAGACTCTAGTAGTATGTATAATCAAAA 4200
Y T N N D D Q H S G C T C N Q G H N P G
TTATACTAACAATGATGACCAGCATTCCGGCTGCACATGTAACCAAGGGCATAACCCTGG 4260
C T C N Q R Y N R * IR IR
CTGTACATGTAATCAAAGATATAACCGTTAACGATTCTAAATAAGAATTAAAATCATTGC 4320
GAAAAATAAAAACCAACTCACAAAATCTATTGCCTATCATAACATAAGCTTTACAAATAA 4380
CTGACATATTCTAGAAGCGGTCTCCTTAATTCTAAAATAAGGAGATCCTTTTCGTTTCCA 4440
CAATATCGATTAATGAAAATACTCCTTTATAGAACGATTTAGGCTGATTGGATTTGAATG 4500
TTGTTGAATCGATGATGAATCTTGAATAGAAATTTGGTATTCAAACTTCTGACGAGGTAT 4560
CTG 4563

Fig. 31. Nucleotide sequence and deduced amino acid sequence of the *cry56Ba1* and *cry39orf2* gene. The *cry56Ba1* gene is 1,980 bp in length and codes for a polypeptide of 659 amino acids. The *cry39orf2* gene is 1,689 bp in length and codes for a polypeptide of 562 amino acids. The potential -35 and -10 boxes and a putative ribosome-binding site (RBS) are marked. The stop codon is marked with asterisks. Five conserved sequence blocks (blocks 1 to 5) are shadowed. Terminal inverted repeats (IR) are indicated below the arrow.

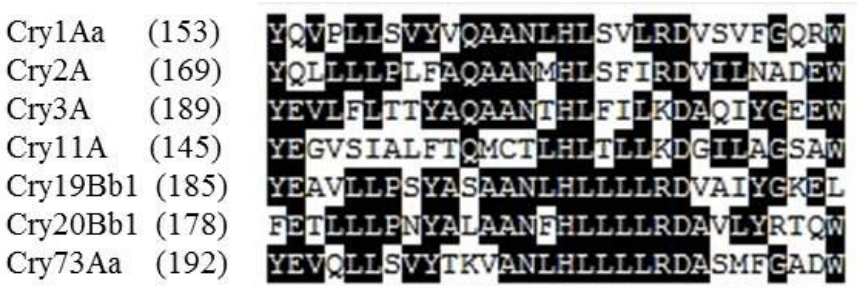
3.2 Analysis of the deduced amino acid sequence of the novel *cry* genes

A Clustal W comparison of the novel proteins from *mogi* with other known Cry protein sequences (http://www.biols.susx.ac.uk/Home/Neil-Crickmore/Bt/) helped to identify the characteristic Cry conserved blocks predicted by Schnepf *et al*. (1998). All of these *cry* genes (*cry19Bb1, cry73Aa, cry20Bb1, cry27Ab1, cry4Aa* and *cry56Ba1*) contained the 5 conserved amino acid residue blocks (Fig. 32 and Fig. 33, block 1 to block 5) that are present in almost all Cry proteins (Schnepf *et al.,* 1998), but does not contain the carboxyl-terminal half of the typical 130 kDa-type crystal proteins (Boonserm *et al.,* 2005; Galitsky *et al.,* 2001; Grochulski *et al*., 1995).

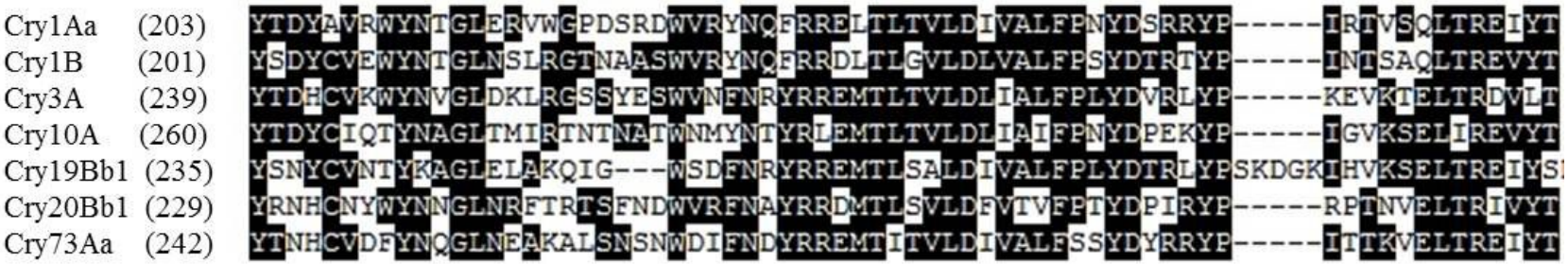
3.3 Transcription level analysis of *cry* genes in *B. thuringiensis* subsp. *mogi*

To confirm the transcription profiles of the selected *cry* genes in *B. thuringiensis* subsp. *mogi* strain, RT-PCR and qPCR were carried out as previous described. The transcription levels of *cry* genes were compared to those of the 16S rRNA gene at each time point. The results suggested that all of these *cry* genes were successfully transcribed in *B. thuringiensis mogi* strain (Fig. 34) in different expression time with different maximum levels (Fig. 35). *B. thuringiensis* subsp. m*ogi* contained *cry27Ab1, cry19Bb1*, *cry20Bb1*, *cry56Ba1*, *cry4Aa* and *cry73Aa* genes, with *cry27Ab1* being transcribed at much higher level and *cry4Aa* being expressed at a relative lower level.

Block 1

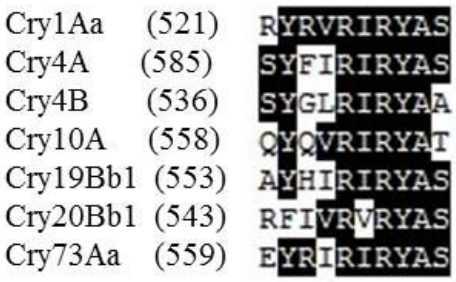


Block 2



Block3

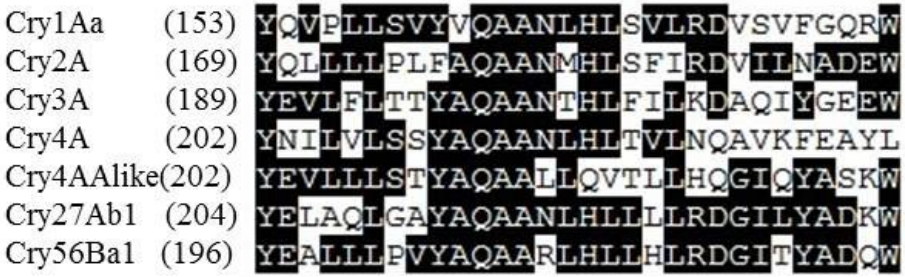
Block 4



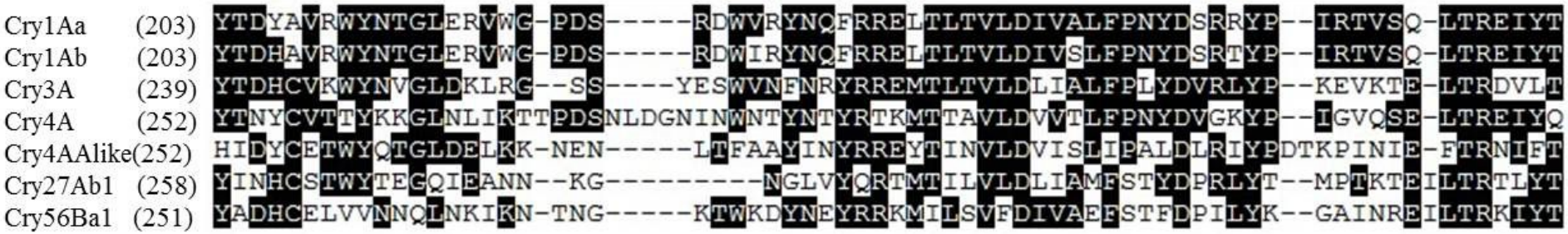
Block 5

Fig. 32. Comparison of the deduced amino acid sequence of Cry19Bb1, Cry20Bb1, Cry73Aa with other Cry proteins in five conserved blocks.

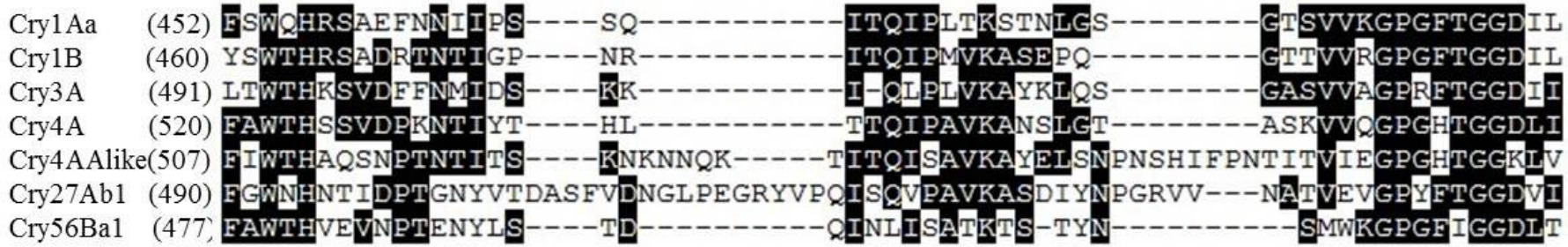
Block 1



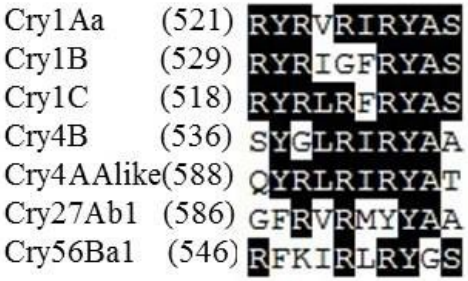
Block 2



Block3



Block 4



Block 5

Fig. 33. Comparison of the deduced amino acid sequence of Cry4Aa, Cry27Ab1, Cry56Ba1 with other Cry proteins in five conserved blocks.

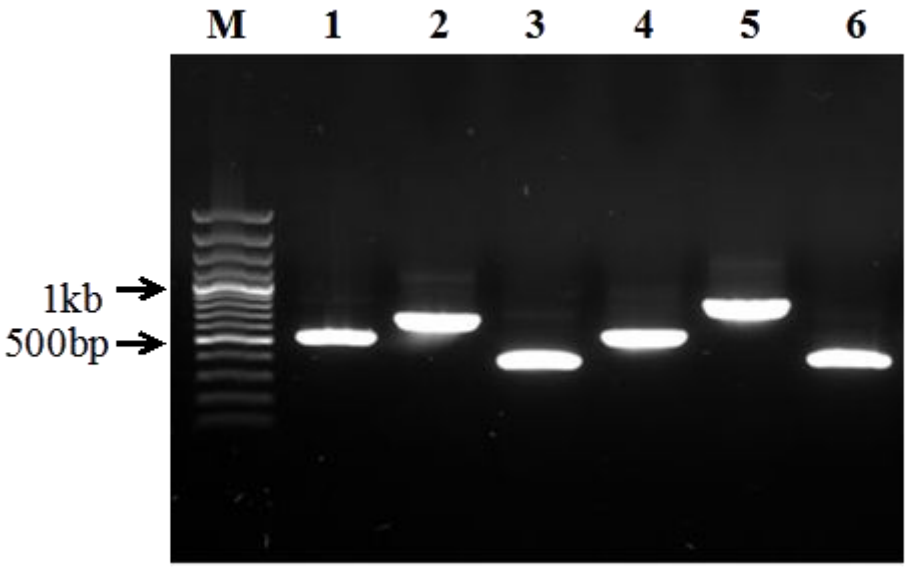


Fig. 34. Agarose gel electrophoresis analysis of *cry* genes mRNA obtained by RT-PCR from *B. thuringiensis* subsp. *mogi* in the sporulation stage. Lane M, Gene Ruler^{TM} 100 bp DNA ladder; lanes 1, *cry27Ab1*; 2, *cry19Bb1*; 3, *cry20Bb1*; 4, *cry56Ba1*; 5, *cry4Aa* and 6, *cry73Aa.*

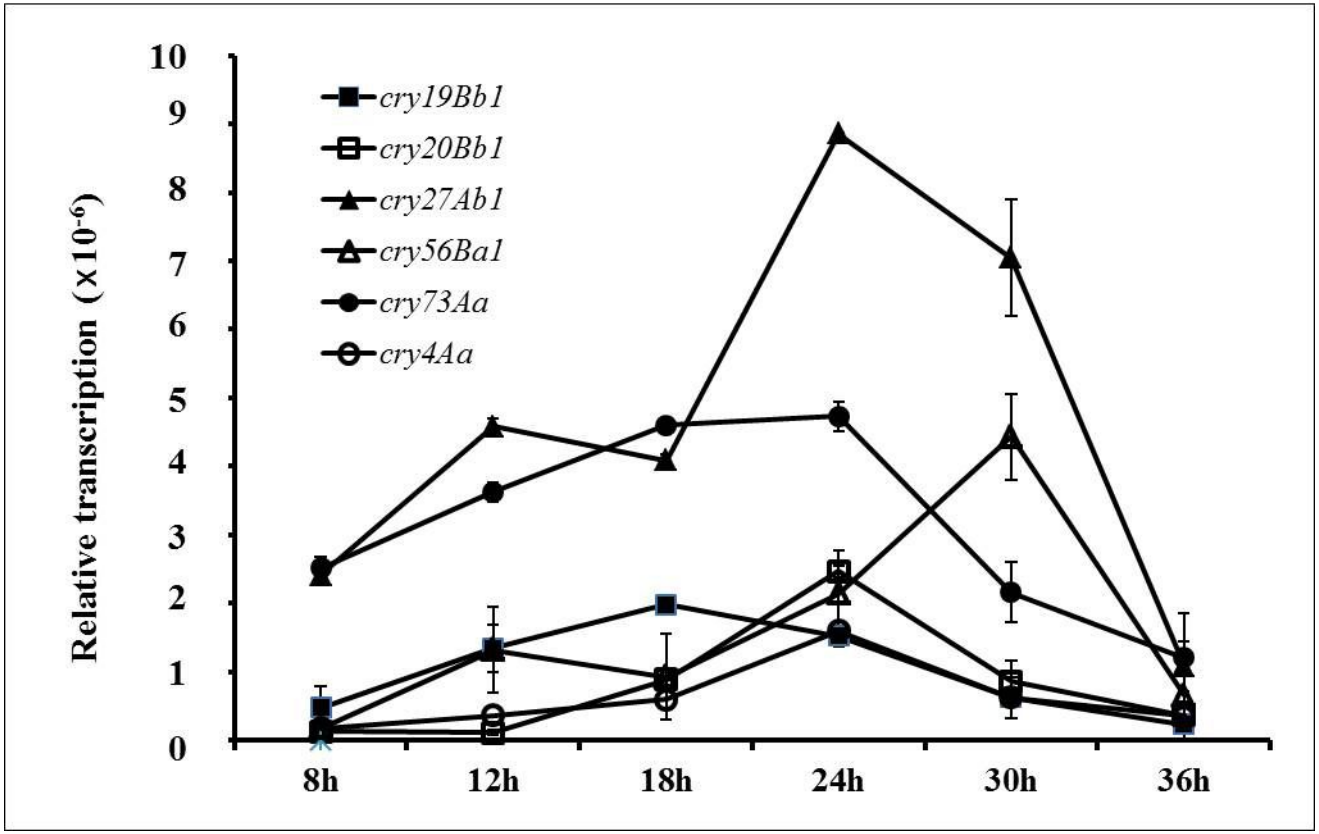


Fig. 35. Transcription-level analysis of the 6 *cry* genes from *B. thuringiensis* subsp. *mogi* were analyzed. The error bars indicate standard deviations.

3.4 Cloning and expression analysis of mosquitocidal *cry* genes in Cry$^-$B

The PCR products (Fig. 36) were cloned into pHT1K (Fig. 22B) and their DNA sequences were confirmed. To investigate the interaction of insecticidal *cry* genes in acrystalliferous *B. thuringiensis*, the recombinant plasmids containing different *cry* genes were introduced into Cry$^-$B by electroporation separately. The expression of these genes were analyzed by SDS-PAGE (Fig. 37).

Cry19Bb1, Cry73Aa with Cry40ORF2, Cry27Ab1 and Cry4Aa, which expressed under their original promoter, were failed to detect on the SDS-PAGE (Fig. 37, lane 2, 3, 5, 6). Meanwhile, there were no crystals visualized in the phase contrast micrographs (Fig. 38, panel 2, 3, 5, 6) either.

The production of Cry20Bb1 by transformant harboring recombinant pHT1K -20Bb1 in Cry$^-$B was shown as the 50 kDa and 30 kDa bands in lane 4 (Fig. 37). The 50- and 30-kDa proteins were confirmed to be the degraded products of the intact 82-kDa Cry20Bb1 protein by N-terminal sequencing. The N-terminal sequences of the 50- and 30-kDa proteins were LLVHV and NVNLQ (Fig. 28), respectively, showing that cleavage occurs between Asp-98 and Leu-99 and between Thr-215 and Asn-216 of the intact 82-kDa Cry20Bb1 protein. This cleavage pattern indicates that both the 50- and 30-kDa proteins have truncated domain I and domain III that are required for insecticidal activity (Chen *et al*., 1993; 1995). The production of Cry56Ba1 with Cry39ORF2 was also detected as the 70 kDa band in lane 7 (Fig. 37).

Transformations crystals of Cry20Bb1 and Cry56Ba1 were visible when phase-contrast microscopy was used, and shown an much smaller inclusion compare with wild type strain (Fig. 38, panel 4 and 7). TEM also demonstrated this results (Fig. 39).

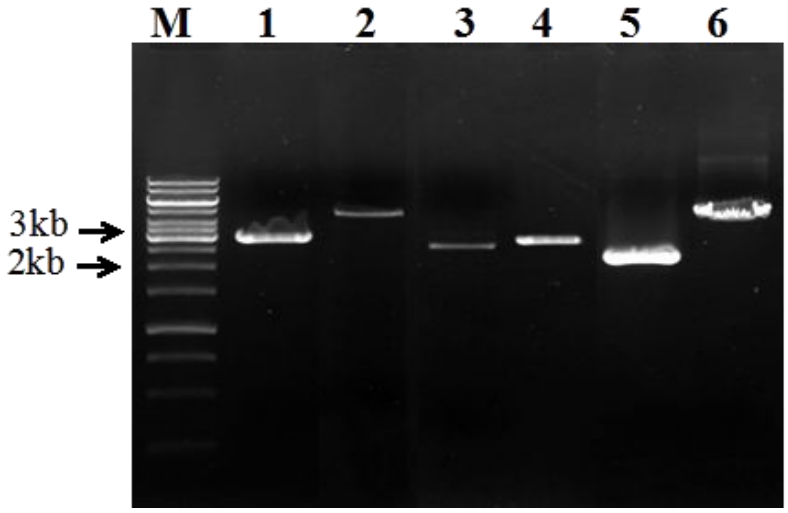


Fig. 36. Agarose gel electrophoresis analysis of PCR products. Lane M, Gene RulerTM 1 kb DNA ladder; lanes 1, *cry19Bb1*; 2, *cry73Aa+cry40orf2*; 3, *cry20Bb1*; 4, *cry27Ab1*; 5, *cry4Aa* and 6, *cry56Ba1+cry39orf2*.

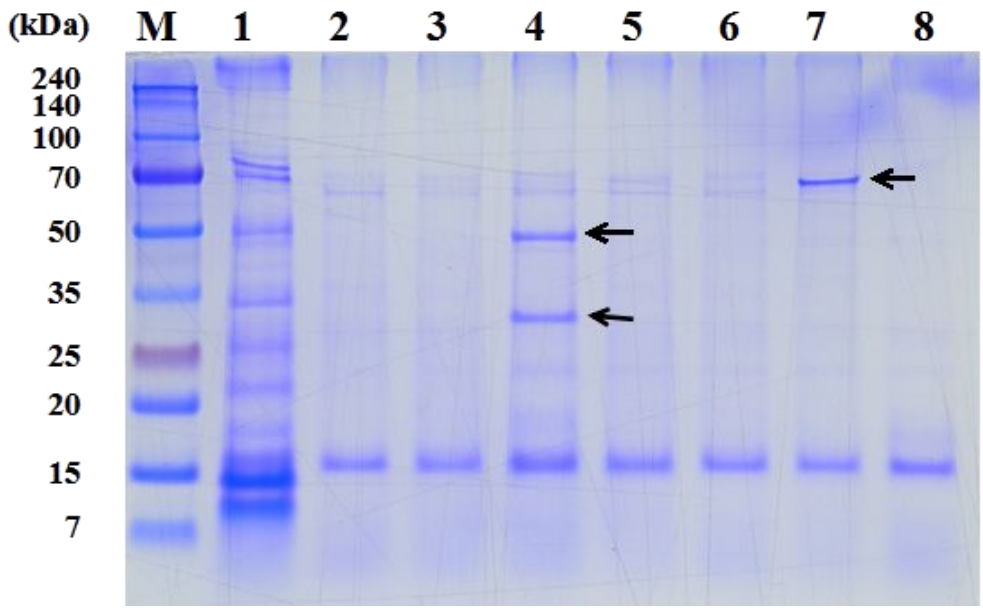


Fig. 37. SDS-PAGE analysis of the recombinant *B. thuringiensis cry⁻B* strains containing different *cry* genes. M, molecular mass marker; lane 1, wild type *mogi* strain; 2, CB/pHT1K-19Bb1; 3, CB/pHT1K-73Aa+40orf2; 4, CB/pHT1K-20Bb1; 5, CB/pHT1K-27Ab1; 6, CB/pHT1K-4Aa; 7, CB/pHT1K-56Ba1+39orf2; 8, wild type Cry⁻B strain.

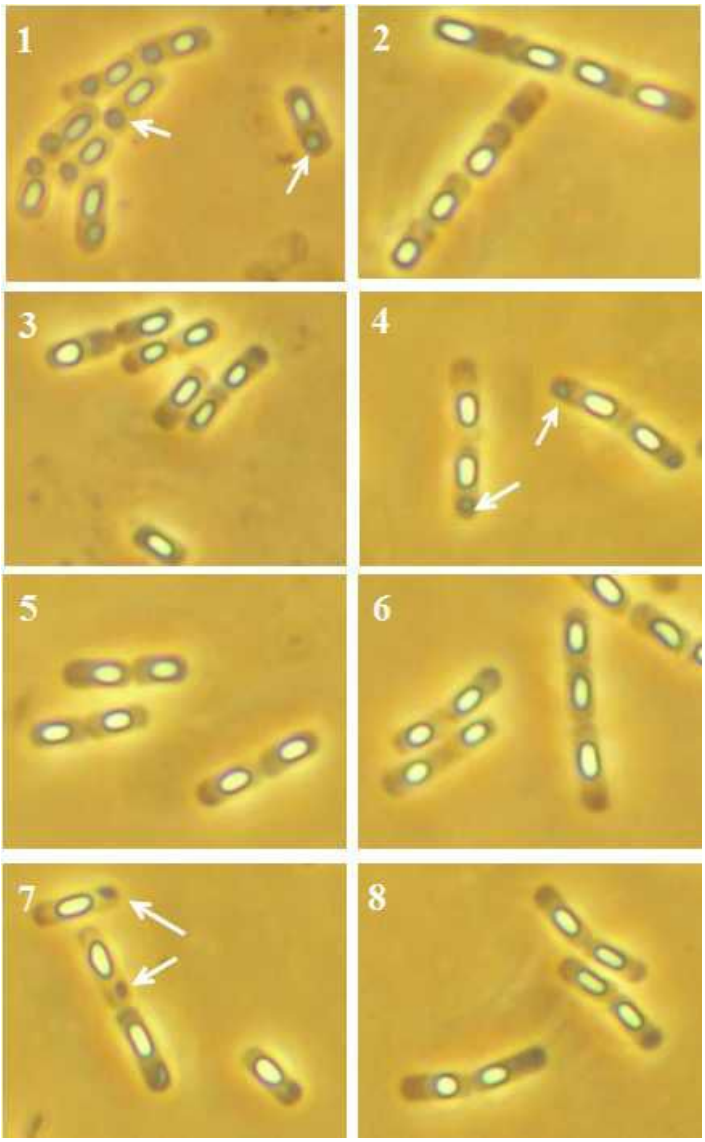


Fig. 38. Phase contrast microscopies of the recombinant *B. thuringiensis* Cry-B strains containing different *cry* genes. 1, wild type *mogi* strain; 2, CB/pHT1K-19Bb1; 3, CB/pHT1K-73Aa+40orf2; 4, CB/pHT1K-20Bb1; 5, CB/pHT1K-27Ab1; 6, CB/pHT1K-4Aa; 7, CB/pHT1K-56Ba1+39orf2; 8, wild type Cry-B strain. Arrows indicate crystal.

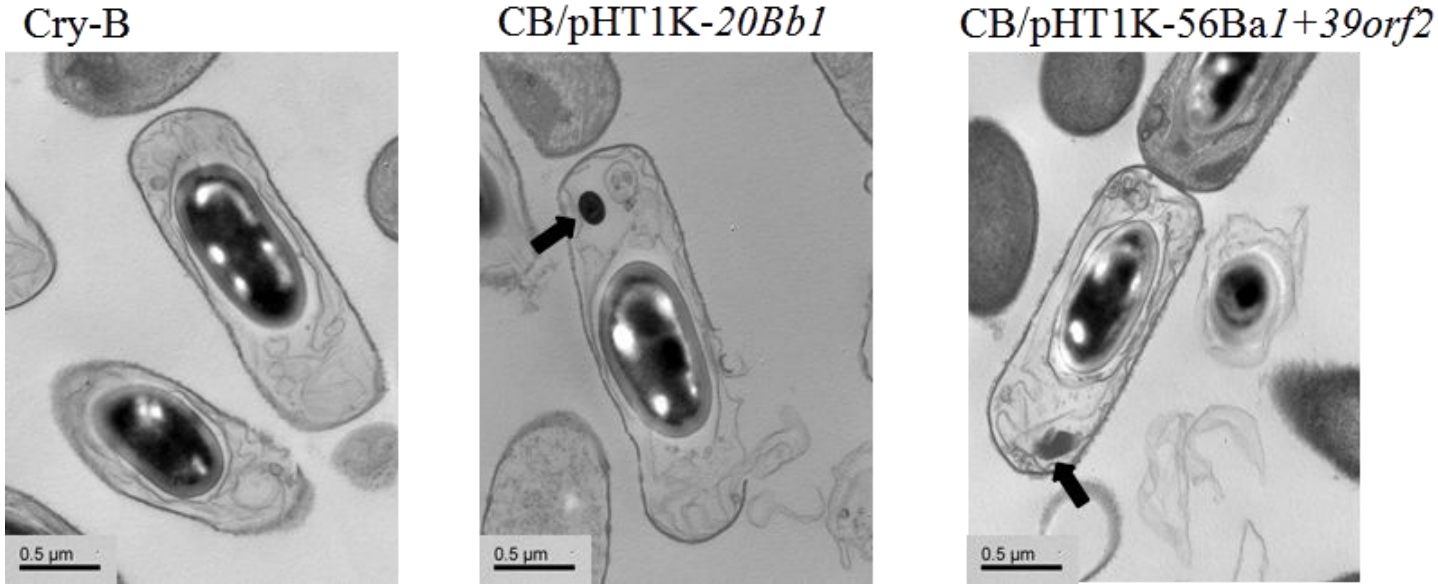


Fig. 39. Transmission electron microscopy of the acrystalliferous *B. thuringiensis* Cry-B strain and the recombinant strains. Magnification is 40,000 ×. Arrows indicate the inclusion.

Table 15. Toxicity of recombinant strains of *B. thuringiensis* Cry-B against *Culex pipiens molestus* and *Cluex pipiens pallens* 4th instar larvae.

Strain	Toxin combination	*Culex pipiens molestus*	*Cluex pipiens pallens*
		LC_{50}[a] (mg/ml)	LC_{50}[a] (mg/ml)
CB/pHT1K-19Bb1	Cry19Bb1	Nt[b]	Nt[b]
CB/pHT1K-73Aa+40orf2	Cry73Aa + Cry40orf2	Nt[b]	Nt[b]
CB/pHT1K-20Bb1	Cry20Bb1	1.17 (0.89-2.06)	1.05 (0.95-1.67)
CB/pHT1K-27Ab1	Cry27Ab1	Nt[b]	Nt[b]
CB/pHT1K-4Aa	Cry4Aa	Nt[b]	Nt[b]
CB/pHT1K-56Ba+39orf2	Cry56Ba1 + Cry39orf2	1.02 (0.91 - 1.54)	1.26 (1.10 -1.42)
Cry-B	no	Nt[b]	Nt[b]

[a]LC_{50}: 50% lethal concentration (in μg) of freeze-dried spore–crystal complex per milliliter after 48 hours. The data are the total of three assays as determined by Probit analysis. [b]Nt, not toxic at 10 mg/ml. [c]FL_{95}: fiducial limits at P=0.95.

3.4 Toxicity of transfromant

The toxicity of wild-type and recombinant strains of *B. thuringiensi* was evaluated against 4th instars of *C. pipiens molestus* and *C. pipiens pallens* (Table 15). Strains CB/pHT1K-19Bb1, CB/pHT1K-73Aa+40orf2, CB/pHT1K-27Ab1 and CB/pHT1K-4Aa, which did not produce visualized inclusions, were not toxic even at 10 mg/ml, while strains CB/pHT1K-20Bb1 and CB/pHT1K-56Ba+39orf2, which produced small but apparently inclusions, showed moderate toxicity, with no significant differences in their median lethal concentrations (LC_{50}).

3.5 Over expression of *cry56Ba1* operon and functional analysis of Cry39ORF2

(i) Cry39ORF2 is required for crystallization of Cry56Ba1

When sporulated cells were examined by phase-contrast microscopy, crystalline inclusions were produced only when both *cry56Ba1* and *cry39orf2* were present in the construct of CB/pHT1K-56Ba+39orf2 (Fig. 23, panel 1; Fig. 40A, panel 1). While there was no crystal observed in the construct of CB/pHT1K-56Ba (Fig. 23, Panel 2; Fig. 40A, Panel 2). Expression of Cry56Ba1 with Cry39ORF2 in construct of CB/pHT1K-56Ba+39orf2 was further confirmed by the SDS-PAGE (Fig. 37 lane7 and Fig. 40B, lane 1).

(ii) Over expression of the *cry56Ba1* operon stabilizes the Cry56Ba1 crystalline inclusions

To determine whether Cry56Ba1 crystals of uniform size and shape could result from over expression of the operon, the *cry56Ba1* (p1KSD-56Ba1; Fig. 23, Pane 4) or *cry56Ba1* and *cry*39orf2 (p1KSD-56Ba1+39orf2; Fig. 23, Pane 3) coding sequences were expressed using the strong chimeric *cyt1A-p*/STAB-SD expression system (Park *et al*., 1998) (Fig. 24 and 25). No inclusion were observed in CB/p1KSD-56Ba that lacked Cry39ORF2 (Fig. 40A, panel 4), whereas crystals of apparently spherical shape were visible in CB/p1KSD-56Ba+39orf2 (Fig. 40A, panel 3), and the size of inclusions in CB/p1KSD-56Ba+39orf2 was much bigger than the one in CB/pHT1K-56Ba+39orf2 (Fig. 40A, panel 1).

SDS-PAGE analysis showed that the amount of Cry56Ba1 synthesized by CB/p1KSD-56Ba+39orf2 (Fig. 40B, lanel 3) was much greater than that produced by CB/pHT1K-56Ba+39orf2 (Fig. 40B, lanel 1), while in recombinant CB/p1KSD-56Ba, which harbored the *cry56Ba1* gene alone, no protein of Cry56Ba1 was detected (Fig. 40B, lanel 4).

The interesting one is, there were big inclusions observed in CB/p1KSD-39orf2 (Fig. 40A, panel 5) while easily degraded and shown a more diffuse protein pattern in SDS-PAGE (Fig. 40B, lanel 5). The amino acid sequence alignment showed that 39ORF2 shared the highest level of identity with the C-terminal region of Cry4Aa (65%), Cry4Ba (65%), Cry7Ba1 (46%), Cry8Aa(46%), and Cry28Aa (46%) (Fig. 41).

Amino acid residues from 165 to 220 of 39ORF2 showed the least identity with C termini of other Cry proteins.

The evaluation of insecticidal activity of different constructs of Cry56Ba1 with Cry39ORF2 proteins against *C. pipiens molestus* and *C. pipiens pallens* 4th instar larvae showed that, the over-expression of recombinant CB/p1KSD-56Ba+39orf2 strain was approximately 14 fold more toxic than CB/pHT1K-56Ba+39orf2 (Table 16). The other 2 strains which contained *cry56Ba1* (CB/pHT1K-56Ba, CB/p1KSD-56Ba) showed no toxicity, even at 10 mg/ml. Interestingly, CB/p1KSD-39orf2, which produced inclusions but quickly degraded also showed no toxicity.

A

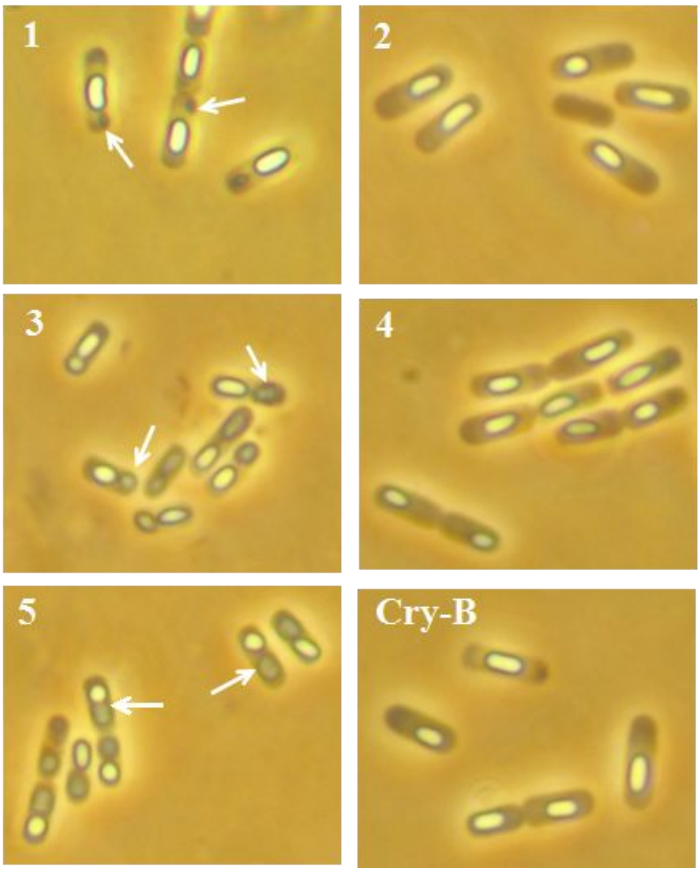


Fig. 40. Over-expression of *cry56Ba1* operon and functional analysis of 39ORF2.

(A) Phase contrast micrographs of the recombinant *B. thuringisnsis* Cry-B strains containing different constructs of *cry56Ba1* genes. 1, CB/pHT1K-56Ba+39orf2; 2, CB/pHT1K-56Ba; 3,CB/p1KSD-56Ba+39orf2; 4, CB/p1KSD-56Ba; 5, CB/p1KSD-39orf2; 6, wild type Cry$^-$B strain. Arrows indicate crystal.

B

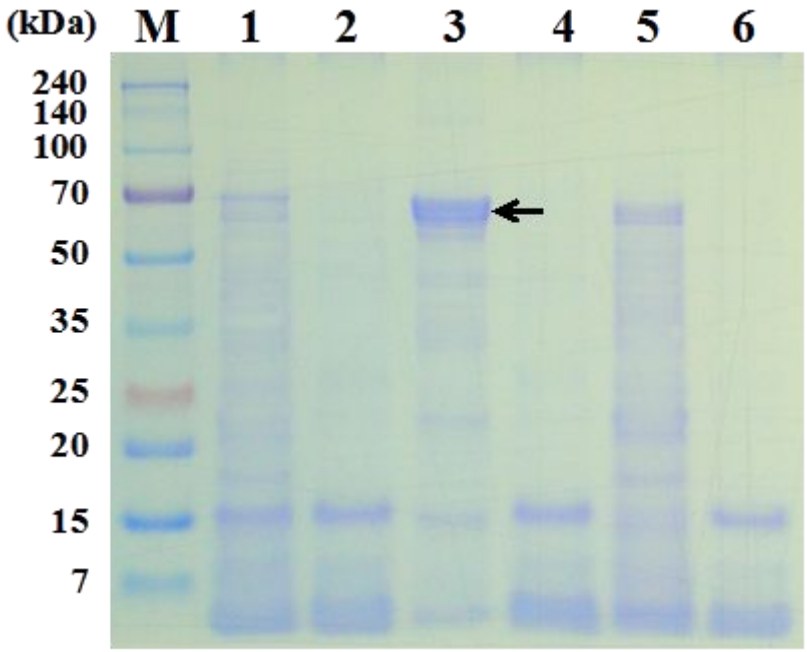


Fig. 40. Over-expression of *cry56Ba1* operon and functional analysis of 39ORF2.

(B) SDS–PAGE analysis of the recombinant *B. thuringiensis* Cry-B strains containing different *cry56Ba1* constructs. M, molecular mass marker; lane 1, CB/pHT1K-56Ba +39orf2; 2, CB/pHT1K-56Ba; 3,CB/p1KSD-56Ba+39orf2; 4, CB/p1KSD-56Ba; 5, CB/p1KSD- 39orf2; 6, wild type Cry$^-$B strain.

Table 16. Toxicity of recombinant strains of *B. thuringiensis* Cry-B against *Culex pipiens molestus* and *Cluex pipiens pallens* 4th instar larvae.

Strain	Toxin combination	*Culex pipiens molestus* LC_{50}[a] (mg/ml)	*Cluex pipiens pallens* LC_{50}[a] (mg/ml)
CB/pHT1K-56Ba+39orf2	Cry56Ba1 + Cry39orf2	1.02 (0.91 - 1.54)	1.26 (1.10 -1.42)
CB/pHT1K-56Ba	Cry56Ba1	Nt[b]	Nt[b]
CB/p1KSD-56Ba+39orf2	Cry56Ba1 + Cry39orf2	0.07 (0.05 -0.11)	0.08 (0.05 -0.12)
CB/p1KSD-56Ba1	Cry56Ba1	Nt[b]	Nt[b]
CB/p1KSD-39orf2	Cry39orf2	Nt[b]	Nt[b]
Cry-B	no	Nt[b]	Nt[b]

[a]LC_{50}: 50% lethal concentration (in μg) of freeze-dried spore–crystal complex per milliliter after 48 hours. The data are the total of three assays as determined by Probit analysis. [b]Nt, not toxic at 10 mg/ml; [c]FL_{95}: fiducial limits at P=0.95.

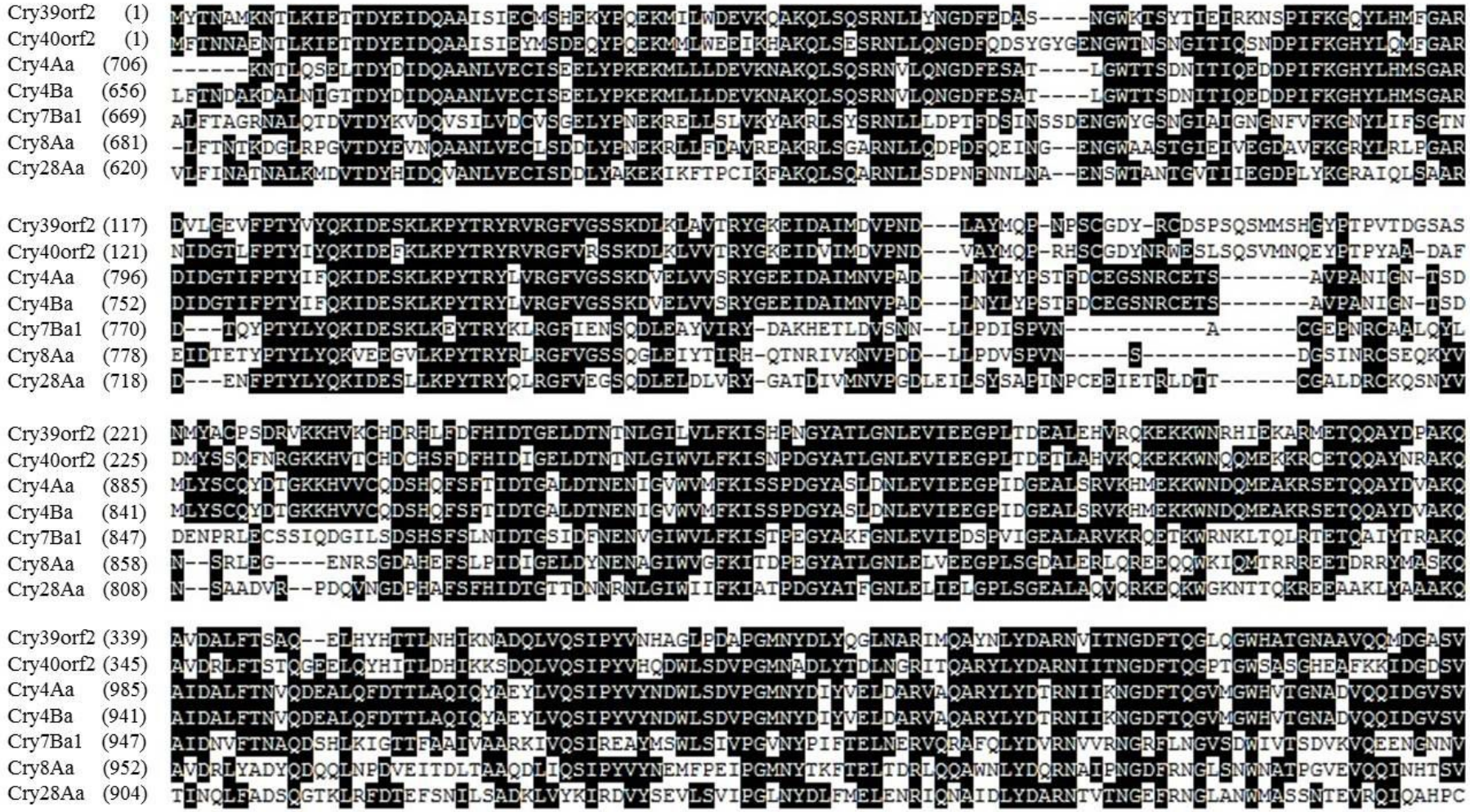

Cry39orf2 (1)
Cry40orf2 (1)
Cry4Aa (706)
Cry4Ba (656)
Cry7Ba1 (669)
Cry8Aa (681)
Cry28Aa (620)
Cry39orf2 (117)
Cry40orf2 (121)
Cry4Aa (796)
Cry4Ba (752)
Cry7Ba1 (770)
Cry8Aa (778)
Cry28Aa (718)
Cry39orf2 (221)
Cry40orf2 (225)
Cry4Aa (885)
Cry4Ba (841)
Cry7Ba1 (847)
Cry8Aa (858)
Cry28Aa (808)
Cry39orf2 (339)
Cry40orf2 (345)
Cry4Aa (985)
Cry4Ba (941)
Cry7Ba1 (947)
Cry8Aa (952)
Cry28Aa (904)

Fig. 41. Alignment of the orf2 amino acid sequence with C terminal regions of selected Cry proteins.

4. Discussion

The crystalline toxic proteins, δ-endotoxins, are predominantly synthesized as large, inactive protoxins that are activated by proteolysis in the insect gut (Gill *et al.*, 1992). The most common type Cry1 protein (about 130 kDa), consist of a N-terminal half containing the toxic portion of the molecule, released after ingestion by insect midgut proteases, and a C-terminal half important to crystallization (Baum and Malvar, 1995; Honee *et al.*, 1991). For example, Cry1, Cry4A, and Cry4B have molecular weights of 130 kDa to 140 kDa and are processed to active 65- to 70-kDa toxins (Gill *et al.*, 1992; Hofte and Whiteley, 1989). While Cry2A, Cry3A, Cry10A, and Cry11A are naturally truncated toxins and have molecular weights that range from 65 kDa to 80 kDa (Höfte and Whiteley, 1989). These proteins are correspond to the N-terminal half of the 130 kDa Cry type. Moreover, proteolytic cleavage at the N and C termini can also process these naturally truncated toxins to active 60 to 65 kDa toxins, as observed in Cry2A and Cry3A (Aronson, 1993).

Attempts to clone and express six *cry* genes (*cry19Bb1, cry20Bb1, cry73Aa* with *cry40orf2, cry27Ab1, cry4Aa* and *cry56Ba1* with *cry39orf2*) from *B.thuringiensis* subsp. *mogi*, four of them were failed to achieve sufficient levels of expression to allow the formation of a parasporal inclusion body as usually occurs with most *B.thuringiensis* endotoxins. Two of them, c*ry20Bb1* and *cry56Ba1 with cry39orf2*

were successfully formed inclusion body in the recombinant *B.thuringiensis* Cry$^-$B strains.

The deduced amino acid sequence of Cry20Bb1 shared a high level of identity (72.6%) with Cry20Aa from *B.thuringiensis* subsp. f*ukuokaensis* (Lee and Gill, 1997). To evaluate the mosquitocidal activity of the Cry20Bb1 protein, attempts were made to obtain purified parasporal inclusions. However, all attempts were unsuccessful since Cry20Bb1 rapidly degraded into tiny inclusions upon cell autolysis. These tiny inclusions contain mostly the smaller 50 and 30 kDa proteins, as determined by SDS-PAGE. Cry20Bb1 was degraded even when protease inhibitors were used during the isolation and purification procedures. The low mosquitocidal activity is not surprising since both the 50 and 30 kDa proteins have truncated domain I and III. Domain I forms pores in the insect midgut cell epithelium and is essential for insecticidal activity (Chen *et al*., 1995). Domain III is increasingly thought to play an essential role in insecticidal activity too (Chen *et al*., 1993).

Since wild-type *B. thuringiensis* subsp. *mogi* contains a number of proteins, each protein not only is a component of the inclusion but may also function to stabilize other proteins. This in part may explain the Cry20Bb1 stability in the wild-type strain but not in transformant Cry$^-$B strain.

Cry56Ba1 encodes a 659 amino acid protein containing only homology blocks 1–5. Homology blocks 6–8 are instead found in the protein Cry39ORF2, which is encoded

by the gene immediately following the intergenic region at the 3'-end of *cry56Ba1*. Eight similar *cry* gene pairs have been described previously in *B. thuringiensis*. The first genes of these eight pairs are *cry10Aa* (Thorne *et al.*, 1986), *cry19Aa* (Rosso and Delécuse, 1997), *cry24Ba* (Ohgushi *et al*., 2005), *cry30Ba* (Ito *et al*., 2006), *cry44Aa* (Ito *et al*., 2006), *cry5Ad* (Lenane *et al.*, 2008), *cry40Aa* and *cry40Ba* (GenBank accession numbers AB074414 and AB112346, respectively). Their upstream reading frames code for the Cry N-terminal domain, and the second frame found approximately 50-100 bp downstream codes for an apparent C-terminal domain that presumably has a function similar to that of ORF2 in protoxin aggregation and crystallization (Barboza-Corona *et al.,* 2012).

The primary genetic factors affecting insecticidal protein synthesis in *B. thuringiensis* are promoters, a 5'mRNA stabilizing sequence and 3' transcriptional termination sequences (Federici *et al*., 2010). The yield increases of Cry56Ba1 obtained with *cyt1AP*/STAB are likely due to higher gene expression resulting from the use of *cyt1A* promoter and especially to greater transcript stability conferred by the STAB-SD sequence (Agaisse and Lereclus, 1996). And the presence of 39orf2, also may stabilize the mRNA, or act as a chaperone to increase the stability of Cry56Ba1.

The toxicity of wild type *B. thuringiensis* subsp. *mogi* strain showed much higher than these recombinant strains. It could be explained by the possibility of synergistic action between different Cry proteins. Notwithstanding the complexity of form and

size, the protoxins are made as inactive protoxins and are activated by proteolysis to toxins. It is not known how these proteins of different size and amino acid sequence fold to generate common protease processing sites. There is virtually no information on the role of glycosylations in protease activation of toxin (Rukmini *et al.*, 2000). The process of activation appears to resemble that of mammalian gut proteases such as pepsinogen and trypsinogen in that a relatively small N-terminal peptide is removed. However, in case of protoxin activation, extensive C-terminal processing is involved and there are no internal cleavages generated within the toxic moiety during activation. It appears that conformational changes occurring during activation are rather subtle, affecting the tertiary structure but not the secondary structure of proteins (Choma and Kaplan, 1990; 1991). The polypeptide of toxic partiality in protoxin when compared to that of active toxin has different thermal unfolding properties.

Table S1. The CDS in pMOGI364 and their annotations.

CDS	Size (aa)	Strand	Annotation	Best hit in databases (GenBank no.)	(% aa identity)
1	473	+	hypothetical protein	hypothetical protein BCG9842_0219 [Bacillus cereus G9842] (YP_002454633.1)	99 in 473 aa
2	190	+	Ser/Thr protein phosphatase	metallophosphoesterase, calcineurin superfamily, putative [Bacillus cereus G9842] (YP_002454634.1)	99 in 190 aa
3	130	+	hypothetical protein	hypothetical protein bthur0007_54260 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111573.1)	95 in 130 aa
4	137	+	hypothetical protein	hypothetical protein IK9_05424 [Bacillus cereus VD166] (ZP_17621097.1)	85 in 136 aa
5	139	+	hypothetical protein	hypothetical protein BCG9842_0223 [Bacillus cereus G9842] (YP_002454637.1)	96 in 139 aa
6	146	+	hypothetical protein	hypothetical protein IK9_05422 [Bacillus cereus VD166] (ZP_17621095.1)	91 in 146 aa
7	147	-	hypothetical protein	hypothetical protein BCG9842_0225 [Bacillus cereus G9842] (YP_002454639.1)	96 in 139 aa
8	215	-	pseudogene		
9	139	+	hypothetical protein	hypothetical protein BCG9842_0227 [Bacillus cereus G9842] (YP_002454641.1)	99 in 139 aa
10	63	+	hypothetical protein	hypothetical protein BCG9842_0228 [Bacillus cereus G9842] (YP_002454642.1)	100 in 63 aa
11	96	+	hypothetical protein	hypothetical protein BCG9842_0229 [Bacillus cereus G9842] (YP_002454643.1)	99 in 96 aa
12	77	+	hypothetical protein	hypothetical protein BCG9842_0230 [Bacillus cereus G9842] (YP_002454644.1)	96 in 77 aa
13	141	+	hypothetical protein	hypothetical protein BCG9842_0231 [Bacillus cereus G9842] (YP_002454645.1)	96 in 141 aa
14	47	+	hypothetical protein	hypothetical protein BCG9842_0232 [Bacillus cereus G9842] (YP_002454646.1)	96 in 47 aa
15	158	+	hypothetical protein	hypothetical protein bthur0007_54170 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111564.1)	87 in 158 aa
16	241	-	S-layer protein	S-layer protein [Bacillus cereus G9842] (YP_002454648.1)	97 in 241 aa
17	288	+	ribosomal protein S1 domain protein	ribosomal protein S1 domain protein [Bacillus cereus G9842] (YP_002454649.1)	100 in 287 aa

18	321	+	hypothetical protein	hypothetical protein BCG9842_0236 [Bacillus cereus G9842] (YP_002454650.1)	100 in 321 aa
19	140	+	hypothetical protein	hypothetical protein BCG9842_0237 [Bacillus cereus G9842] (YP_002454651.1)	99 in 140 aa
20	595	+	hypothetical protein	hypothetical protein BCG9842_0238 [Bacillus cereus G9842] (YP_002454652.1)	97 in 595 aa
21	100	+	hypothetical protein	hypothetical protein BCG9842_0239 [Bacillus cereus G9842] (YP_002454653.1)	99 in 100 aa
22	121	+	hypothetical protein	hypothetical protein BCG9842_0240 [Bacillus cereus G9842] (YP_002454654.1)	98 in 120 aa
23	158	+	hypothetical protein	hypothetical protein BCG9842_0241 [Bacillus cereus G9842] (YP_002454655.1)	100 in 158 aa
24	135	+	single-strand binding protein	single-strand binding protein family [Bacillus cereus G9842] (YP_002454656.1)	96 in 135 aa
25	570	-	hypothetical protein	hypothetical protein BCG9842_0243 [Bacillus cereus G9842] (YP_002454657.1)	99 in 570 aa
26	484	+	hypothetical protein	hypothetical protein BCG9842_0244 [Bacillus cereus G9842] (YP_002454658.1)	99 in 484 aa
27	574	+	RpiR family transcriptional regulator	hypothetical protein BCG9842_0245 [Bacillus cereus G9842] (YP_002454659.1)	99 in 574 aa
28	129	-	hypothetical protein	hypothetical protein BCG9842_0246 [Bacillus cereus G9842] (YP_002454660.1)	99 in 129 aa
29	154	-	hypothetical protein	hypothetical protein BCG9842_0247 [Bacillus cereus G9842] (YP_002454661.1)	98 in 153 aa
30	248	+	hypothetical protein	hypothetical protein BCG9842_0248 [Bacillus cereus G9842] (YP_002454662.1)	97 in 248 aa
31	199	+	hypothetical protein	hypothetical protein BCG9842_0249 [Bacillus cereus G9842] (YP_002454663.1)	98 in 199 aa
32	361	+	hypothetical protein	hypothetical protein BCG9842_0250 [Bacillus cereus G9842] (YP_002454664.1)	99 in 361 aa
33	173	-	hypothetical protein	hypothetical protein BCG9842_0251 [Bacillus cereus G9842] (YP_002454665.1)	98 in 166 aa
34	51	+	hypothetical protein	hypothetical protein BCG9842_0252 [Bacillus cereus G9842] (YP_002454666.1)	100 in51 aa
35	75	+	hypothetical protein	hypothetical protein bthur0001_57990 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149204.1)	69 in 55 aa
36	516	+	hypothetical protein	hypothetical protein BCG9842_0254 [Bacillus cereus G9842] (YP_002454668.1)	89 in 437 aa
37	188	+	hypothetical protein	hypothetical protein bthur0007_59340 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112055.1)	96 in 187 aa
38	1116	+	DEAD/DEAH box helicase DEAD-like helicase	helicase conserved C- domain protein [Bacillus cereus G9842] (YP_002454670.1)	99 in 1116 aa

39	287	+	hypothetical protein	hypothetical protein BCG9842_0005 [Bacillus cereus G9842] (YP_002454671.1)	100 in 287 aa
40	332	+	acetyltransferase	YdjC [Bacillus cereus G9842] (YP_002454672.1)	99 in 332 aa
41	450	+	Sporulation kinase	multi-sensor signal transduction histidine kinase, putative [Bacillus cereus G9842] (YP_002454673.1)	99 in 450 aa
42	191	+	NarL family DNA-binding response regulator	Two-component response regulator YhcZ, putative [Bacillus cereus G9842] (YP_002454674.1)	99 in 191 aa
43	60	+	HTH-type transcriptional regulator SinR	HTH-type transcriptional regulator SinR [Bacillus cereus G9842] (YP_002454675.1)	98 in 60 aa
44	391	+	hypothetical protein	hypothetical protein BCG9842_0010 [Bacillus cereus G9842] (YP_002454676.1)	98 in 391 aa
45	50	+	hypothetical protein	hypothetical protein BCG9842_0011 [Bacillus cereus G9842] (YP_002454677.1)	96 in 50 aa
46	171	+	hypothetical protein	hypothetical protein BCG9842_0012 [Bacillus cereus G9842] (YP_002454678.1)	99 in 169 aa
47	64	+	hypothetical protein	hypothetical protein BCG9842_0012 [Bacillus cereus G9842] (YP_002454678.1)	98 in 64 aa
48	179	+	hypothetical protein	hypothetical protein bthur0001_54190 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148849.1)	98 in 197 aa
49	191	+	hypothetical protein	hypothetical protein BCG9842_0014 [Bacillus cereus G9842] (YP_002454680.1)	97 in 191 aa
50	548	+	DNA polymerase III subunit gamma/tau	DNA polymerase III subunit gamma/tau [Bacillus cereus G9842] (YP_002454681.1)	99 in 548 aa
51	55	-	hypothetical protein	hypothetical protein bthur0007_59220 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112043.1)	100 in 55 aa
52	121	-	hypothetical protein	YtvB [Bacillus cereus G9842] (YP_002454683.1)	99 in 121 aa
53	152	+	hypothetical protein	hypothetical protein bthur0007_59200 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112041.1)	97 in 152 aa
54	199	+	hypothetical protein	hypothetical protein IK9_05618 [Bacillus cereus VD166] (ZP_17621291.1)	96 in 199 aa
55	51	+	hypothetical protein	hypothetical protein bthur0007_59180 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112039.1)	100 in 51 aa
56	663	+	phosphoadenosine phosphosulfate reductase	hypothetical protein BCG9842_0022 [Bacillus cereus G9842] (YP_002454688.1)	98 in 388 aa

57	388	+	sulfurtransferase DndC	hypothetical protein BCG9842_0022 [Bacillus cereus G9842] (YP_002454688.1)	98 in 388 aa
58	174	-	signal peptidase I	signal peptidase I [Bacillus cereus G9842] (YP_002454689.1)	93 in 174 aa
59	163	+	oligopeptide ABC transporter ATP-binding protein	hypothetical protein bthur0007_59140 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112035.1)	96 in 162 aa
60	122	+	cell division protein SepF	hypothetical protein BCG9842_0025 [Bacillus cereus G9842] (YP_002454691.1)	95 in 122 aa
61	624	+	HNH endonuclease	hypothetical protein BCG9842_0026 [Bacillus cereus G9842] (YP_002454692.1)	81 in 344 aa
62	219	+	group-specific protein	hypothetical protein IK9_05610 [Bacillus cereus VD166] (ZP_17621283.1)	94 in 224 aa
63	415	+	glutathionylspermidine synthase	Glutathionylspermidine synthase [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148930.1)	97 in 415 aa
64	75	-	hypothetical protein	hypothetical protein IK9_05608 [Bacillus cereus VD166] (ZP_17621281.1)	93 in 45 aa
65	157	+	hypothetical protein	hypothetical protein BCG9842_0029 [Bacillus cereus G9842] (YP_002454695.1)	99 in 157 aa
66	214	+	thermonuclease family protein	thermonuclease family protein [Bacillus cereus G9842] (YP_002454696.1)	99 in 214 aa
67	89	+	actin binding protein	hypothetical protein IK9_05605 [Bacillus cereus VD166] (ZP_17621278.1)	100 in 89 aa
68	201	+	hypothetical protein	hypothetical protein BCG9842_0032 [Bacillus cereus G9842] (YP_002454698.1)	100 in 200 aa
69	117	+	hypothetical protein	hypothetical protein BCG9842_0033 [Bacillus cereus G9842] (YP_002454699.1)	98 in 117 aa
70	214	-	signal peptidase I	signal peptidase I [Bacillus cereus VD166] (ZP_17621275.1)	99 in 214 aa
71	286	+	RNA polymerase sigma factor RpoD	RNA polymerase sigma factor, sigma-70 family [Bacillus cereus G9842] (YP_002454701.1)	97 in 284 aa
72	127	+	hypothetical protein	hypothetical protein bthur0001_55210 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148937.1)	94 in 127 aa
73	84	+	recombination protein RecR	hypothetical protein IK9_05599 [Bacillus cereus VD166] (ZP_17621272.1)	92 in 84 aa
74	152	+	hypothetical protein	hypothetical protein bthur0001_55230 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148939.1)	94 in 152 aa
75	43	+	hypothetical protein	hypothetical protein BCG9842_0039 [Bacillus cereus G9842] (YP_002454705.1)	98 in 43 aa

76	81	+	hypothetical protein	hypothetical protein BCG9842_0039 [Bacillus cereus G9842] (YP_002454705.1)	99 in 81 aa
77	90	+	hypothetical protein	hypothetical protein bthur0007_58950 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112016.1)	98 in 90 aa
78	313	+	hypothetical protein	hypothetical protein bthur0007_58940 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112015.1)	98 in 313 aa
79	583	+	hypothetical protein	hypothetical protein BCG9842_0042 [Bacillus cereus G9842] (YP_002454708.1)	99 in 583 aa
80	323	+	DNA polymerase III subunit delta	DNA polymerase III, delta subunit [Bacillus cereus G9842] (YP_002454709.1)	100 in 232 aa
81	179	+	hypothetical protein	hypothetical protein BCG9842_0044 [Bacillus cereus G9842] (YP_002454710.1)	99 in 179 aa
82	57	+	pseudogene		
83	250	-	transposase IstB	hypothetical protein IK5_00386 [Bacillus cereus VD154] (ZP_17603283.1)	98 in 250 aa
84	431	-	transposase IstA	transposase for insertion sequence element IS232 [Bacillus cereus VD154] (ZP_17609012.1)	91 in 431 aa
85	188	+	pseudogene		
86	83	+	host factor Hfq	host factor Hfq [Bacillus cereus G9842] (YP_002454712.1)	100 in 83 aa
87	133	+	hypothetical protein	hypothetical protein BCG9842_0047 [Bacillus cereus G9842] (YP_002454713.1)	99 in 133 aa
88	214	+	hypothetical protein	hypothetical protein BCG9842_0048 [Bacillus cereus G9842] (YP_002454714.1)	98 in 214 aa
89	152	+	hypothetical protein	hypothetical protein BCG9842_0049 [Bacillus cereus G9842] (YP_002454715.1)	98 in 152 aa
90	37	+	hypothetical protein	hypothetical protein BCG9842_0050 [Bacillus cereus G9842] (YP_002454716.1)	100 in 37 aa
91	96	+	hypothetical protein	hypothetical protein BCG9842_0051 [Bacillus cereus G9842] (YP_002454717.1)	96 in 96 aa
92	92	+	transcriptional repressor PagR	transcriptional regulator, ArsR family [Bacillus cereus G9842] (YP_002454718.1)	100 in 92 aa
93	160	+	hypothetical protein	hypothetical protein BCG9842_0053 [Bacillus cereus G9842] (YP_002454719.1)	99 in 160 aa
94	124	+	hypothetical protein	hypothetical protein BCG9842_0054 [Bacillus cereus G9842] (YP_002454720.1)	98 in 124 aa
95	97	+	hypothetical protein	hypothetical protein BCG9842_0055 [Bacillus cereus G9842] (YP_002454721.1)	100 in 97 aa
96	79	+	hypothetical protein	hypothetical protein IK9_05579 [Bacillus cereus VD166] (ZP_17621252.1)	88 in77 aa

97	76	+	hypothetical protein	hypothetical protein IK9_05578 [Bacillus cereus VD166] (ZP_17621251.1)	89 in 76 aa
98	380	-	integrase-recombinase	hypothetical protein IK9_05577 [Bacillus cereus VD166] (ZP_17621250.1)	94 in 380 aa
99	276	+	protein translocase subunit secA	protein translocase subunit secA [Bacillus cereus VD166] (ZP_17621249.1)	91 in 276 aa
100	101	+	hypothetical protein	hypothetical protein IK9_05575 [Bacillus cereus VD166] (ZP_17621248.1)	90 in 101 aa
101	92	+	transition state regulatory protein AbrB	Transition state regulatory protein AbrB [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111996.1)	97 in 92.aa
102	72	+	membrane protein	membrane protein [Planococcus antarcticus DSM 14505] (ZP_10206699.1)	70 in 71 aa
103	238	+	ABC transporter ATP-binding protein	ABC transporter, ATP-binding protein [Planococcus antarcticus DSM 14505] (ZP_10206700.1)	74 in 238 aa
104	457	+	membrane protein	hypothetical protein A1A1_01488 [Planococcus antarcticus DSM 14505] (ZP_10206701.1)	64 in 458 aa
105	91	+	hypothetical protein	hypothetical protein A1A1_01493 [Planococcus antarcticus DSM 14505] (ZP_10206702.1)	69 in 85 aa
106	125	+	membrane protein	membrane protein [Planococcus antarcticus DSM 14505] (ZP_10206703.1)	54 in 125 aa
107	145	+	disulfide formation protein C	hypothetical protein IE5_05407 [Bacillus cereus BAG3X2-2] (ZP_17404749.1)	67 in 141 aa
108	324	+	DNA-methyltransferase	DNA-methyltransferase [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148963.1)	90 in 330 aa
109	92	+	membrane protein	hypothetical protein IK9_05571 [Bacillus cereus VD166] (ZP_17621244.1)	93 in 92 aa
110	228	+	hypothetical protein	hypothetical protein ICU_04726 [Bacillus cereus BAG2X1-1] (ZP_17376233.1)	93 in 228 aa
111	71	+	Small, acid-soluble spore protein C	Small, acid-soluble spore protein C [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111991.1)	97 in 71 aa
112	140	+	hypothetical protein	hypothetical protein BCG9842_0066 [Bacillus cereus G9842] (YP_002454732.1)	89 in 140 aa
113	62	+	hypothetical protein	hypothetical protein bthur0007_58680 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111989.1)	80 in 61 aa

114	190	+	hypothetical protein	conserved hypothetical protein [Bacillus cereus AH1134] (ZP_03233313.1)	80 in 109 aa
115	73	+	hypothetical protein	hypothetical protein II7_05369 [Bacillus cereus MSX-A12] (ZP_17548393.1)	95 in 73 aa
116	55	+	hypothetical protein	hypothetical protein IG1_05794 [Bacillus cereus HD73] (ZP_17484757.1)	78 in 55 aa
117	102	+	hypothetical protein	hypothetical protein bthur0008_37170 [Bacillus thuringiensis serovar berliner ATCC 10792] (ZP_04103634.1)	91 in 101 aa
118	241	+	hypothetical protein	hypothetical protein IC1_05843 [Bacillus cereus VD022] (ZP_17341366.1)	93 in 241 aa
119	72	+	hypothetical protein	hypothetical protein IIE_06160 [Bacillus cereus VD045] (ZP_17566835.1)	91 in 66 aa
120	65	+	hypothetical protein	hypothetical protein bthur0007_58660 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111987.1)	85 in 54 aa
121	150	+	vacuolar protein-sorting -associated protein 36	hypothetical protein bthur0007_58650 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111986.1)	88 in 150 aa
122	326	+	MerR family transcriptional regulator	transcriptional regulator, MerR [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111984.1)	97 in 326 aa
123	306	+	WXG100 family type VII secretion target	hypothetical protein bthur0007_58620 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111983.1)	99 in 306 aa
124	119	+	hypothetical protein	hypothetical protein bthur0007_58610 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1]	100 in 119 aa
125	90	-	DNA-binding protein HU	DNA-binding protein HU [Bacillus cereus VD166] (ZP_17621231.1)	96 in 90 aa
126	81	+	hypothetical protein	hypothetical protein bthur0007_61170 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112230.1)	98 in 81 aa
127	52	-	hypothetical protein	hypothetical protein bthur0007_55320 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111679.1)	82 in 50 aa
128	357	-	response regulator aspartate phosphatase	hypothetical protein bthur0001_55580 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148974.1)	100 in 357 aa

129	147	+	ArsR family transcriptional regulator	hypothetical protein bthur0001_55590 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148975.1)	99 in 147 aa
130	64	+	hypothetical protein	hypothetical protein bthur0001_55600 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148976.1)	100 in 64 aa
131	327	+	hypothetical protein	hypothetical protein bthur0001_55610 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148977.1)	99 in 327 aa
132	133	+	hypothetical protein	hypothetical protein bthur0007_61210 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112234.1)	98 in 133 aa
133	244	+	hypothetical protein	hypothetical protein bthur0001_55630 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148979.1)	99 in 241 aa
134	82	+	hypothetical protein	hypothetical protein IK9_05548 [Bacillus cereus VD166] (ZP_17621221.1)	100 in 82 aa
135	60	-	hypothetical protein	hypothetical protein BCG9842_0085 [Bacillus cereus G9842] (YP_002454751.1)	100 in 60 aa
136	185	+	histidinol-phosphate phosphatase domain-containing protein	histidinol-phosphate phosphatase domain-containing protein [Bacillus cereus VD166] (ZP_17621219.1)	95 in 185 aa
137	141	+	thioredoxin	hypothetical protein BCG9842_0087 [Bacillus cereus G9842] (YP_002454753.1)	97 in 141 aa
138	173	+	hypothetical protein	hypothetical protein bthur0001_55670 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148983.1)	98 in 173 aa
139	152	+	hypothetical protein	hypothetical protein BCG9842_0089 [Bacillus cereus G9842] (YP_002454755.1)	98 in 152 aa
140	157	+	hypothetical protein	hypothetical protein BCG9842_0090 [Bacillus cereus G9842] (YP_002454756.1)	99 in 157 aa
141	98	+	hypothetical protein	hypothetical protein BCG9842_0091 [Bacillus cereus G9842] (YP_002454757.1)	98 in 98 aa
142	273	+	UvrD/REP helicase	helicase, UvrD/Rep family [Bacillus cereus G9842] (YP_002454758.1)	99 in 270 aa
143	612	+	group II intron-encoded protein LtrA	group II intron-encoded protein LtrA [Bacillus cereus 03BB108] (ZP_03114863.1)	89 in 612 aa
144	377	+	ATP-dependent DNA helicase PcrA	ATP-dependent DNA helicase PcrA [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112242.1)	96 in 377 aa

145	326	+	DNA polymerase III subunit delta'	DNA polymerase III subunit delta' [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112243.1)	98 in 326 aa
146	87	+	hypothetical protein	hypothetical protein BCG9842_0094 [Bacillus cereus G9842] (YP_002454760.1)	61 in 94 aa
147	312	+	RimK domain-containing protein ATP-grasp	hypothetical protein SchaN1_13993 [Streptomyces chartreusis NRRL 12338] (ZP_09954292.1)	31 in 314 aa
148	327	-	intracellular serine protease	Intracellular serine protease [Bacillus cereus 172560W] (ZP_04309338.1)	65 in 327 aa
149	305	-	carboxypeptidase domain-containing protein	hypothetical protein PpisJ2_04638 [pseudogenealteromonas piscicida JCM 20779] (ZP_10288245.1)	31 in 275 aa
150	75	+	Rev-Erb beta 2	hypothetical protein bthur0001_56290 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149037.1)	99 in 75 aa
151	151	+	hypothetical protein	hypothetical protein bthur0001_56280 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149036.1)	98 in 151 aa
152	121	+	hypothetical protein	hypothetical protein bthur0001_56270 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149035.1)	97 in 121 aa
153	581	+	hypothetical protein	hypothetical protein bthur0007_61360 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112249.1)	94 in 581 aa
154	81	+	hypothetical protein	hypothetical protein IK9_05532 [Bacillus cereus VD166] (ZP_17621205.1)	94 in 81 aa
155	438	+	DNA (cytosine-5-) -methyltransferase	DNA (cytosine-5-)-methyltransferase [Bacillus cereus VD166] (ZP_17621204.1)	99 in 438 aa
156	209	+	thermonuclease	thermonuclease [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149031.1)	99 in 209 aa
157	299	+	foldase protein PrsA	foldase protein PrsA [Bacillus cereus G9842] (YP_002454772.1)	98 in 299 aa
158	114	+	hypothetical protein	hypothetical protein bthur0007_61430 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112256.1)	94 in 114 aa

159	58	+	hypothetical protein	hypothetical protein bthur0007_61440 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112257.1)	100 in 58 aa
160	156	+	hypothetical protein	hypothetical protein bthur0007_61450 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112258.1)	96 in 156 aa
161	131	+	hypothetical protein	hypothetical protein bthur0007_61460 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112259.1)	92 in 131 aa
162	87	+	hypothetical protein	hypothetical protein bthur0007_61470 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112260.1)	97 in 87 aa
163	132	+	hypothetical protein	hypothetical protein bthur0007_61490 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112262.1)	87 in 132 aa
164	243	+	hypothetical protein	hypothetical protein IE1_05529 [Bacillus cereus BAG3O-2] (ZP_17393345.1)	79 in 243 aa
165	135	+	hypothetical protein	hypothetical protein IK9_05524 [Bacillus cereus VD166] (ZP_17621197.1)	89 in 135 aa
166	109	+	hypothetical protein	hypothetical protein IK9_05522 [Bacillus cereus VD166] (ZP_17621195.1)	95 in 109 aa
167	167	-	hypothetical protein	hypothetical protein bthur0007_61520 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112265.1)	99 in 167 aa
168	463	-	Type II DNA-methyltransferase	Type II DNA-methyltransferase [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112266.1)	98 in 463 aa
169	116	+	hypothetical protein	hypothetical protein bthur0001_56920 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149100.1)	95 in 116 aa
170	144	+	hypothetical protein	hypothetical protein bthur0001_56910 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149099.1)	93 in 144 aa
171	119	+	hypothetical protein	hypothetical protein BCG9842_0121 [Bacillus cereus G9842] (YP_002454787.1)	93 in 119
172	337	-	primosomal protein DnaI	primosomal protein DnaI [Bacillus cereus G9842] (YP_002454788.1)	98 in 337 aa
173	86	-	hypothetical protein	hypothetical protein BCG9842_0123 [Bacillus cereus G9842] (YP_002454789.1)	99 in 86 aa

174	65	-	hypothetical protein	hypothetical protein IK9_05515 [Bacillus cereus VD166] (ZP_17621188.1)	94 in 65 aa
175	79	+	hypothetical protein	hypothetical protein bthur0001_56860 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (WP_000043819.1)	94 in 79 aa
176	62	+	RNA chaperone Hfq	Host factor-I protein [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149093.1)	98 in 62 aa
177	459	+	hypothetical protein	hypothetical protein bthur0007_61590 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04112272.1)	76 in 459 aa
178	99	+	ArsR family transcriptional regulator	transcriptional regulator, ArsR family [Bacillus cereus G9842] (YP_002454793.1)	99 in 99 aa
179	133	+	hypothetical protein	hypothetical protein bthur0001_56820 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149090.1)	99 in 133 aa
180	72	+	hypothetical protein	ypothetical protein IK9_05507 [Bacillus cereus VD166] (ZP_17621180.1)	100 in 72 aa
181	96	+	hypothetical protein	hypothetical protein BCG9842_0135 [Bacillus cereus G9842] (YP_002454801.1)	94 in 96 aa
182	361	+	integrase/recombinase, phage integrase family protein	integrase/recombinase, phage integrase family protein [Bacillus cereus G9842] (YP_002454802.1)	99 in 361 aa
183	174	-	hypothetical protein	hypothetical protein bthur0001_56800 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149088.1)	98 in 174 aa
184	239	-	hypothetical protein	hypothetical protein BCG9842_0138 [Bacillus cereus G9842] (YP_002454804.1)	96 in 239 aa
185	115	-	ribosome biogenesis GTPase rsgA	hypothetical protein BCG9842_0139 [Bacillus cereus G9842] (YP_002454805.1)	96 in 114 aa
186	420	-	DNA-damage repair protein	DNA-damage repair protein [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04111654.1)	92 in 421 aa
187	394	+	RES domain-containing protein	hypothetical protein IK9_05501 [Bacillus cereus VD166] (ZP_17621174.1)	96 in 392 aa
188	118	-	hypothetical protein	TrsE, putative [Bacillus cereus G9842] (YP_002454807.1)	56 in 118 aa
189	60	-	hypothetical protein	hypothetical protein BCG9842_0142 [Bacillus cereus G9842] (YP_002454808.1)	97 in 60 aa
190	167	-	hypothetical protein	hypothetical protein BCG9842_0143 [Bacillus cereus G9842] (YP_002454809.1)	81 in 167 aa

191	131	-	hypothetical protein	hypothetical protein BCG9842_0144 [Bacillus cereus G9842] (YP_002454810.1)	94 in 131 aa
192	73	-	hypothetical protein	hypothetical protein bthur0001_56730 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149081.1)	96 in 73 aa
193	137	-	hypothetical protein	hypothetical protein bthur0013_54750 [Bacillus thuringiensis IBL 200] (ZP_04075141.1)	71 in 128 aa
194	173	-	hypothetical protein	hypothetical protein IGK_05536 [Bacillus cereus HuB4-10] (ZP_17519835.1)	69 in 165 aa
195	61	-	hypothetical protein	hypothetical protein BCG9842_0154 [Bacillus cereus G9842] (YP_002454820.1)	83 in 60 aa
196	182	-	hypothetical protein	hypothetical protein IK9_05488 [Bacillus cereus VD166] (ZP_17621161.1)	88 in 180 aa
197	465	-	hypothetical protein	hypothetical protein BCG9842_0156 [Bacillus cereus G9842] (YP_002454822.1)	99in 465 aa
198	96	-	hypothetical protein	hypothetical protein BCG9842_0157 [Bacillus cereus G9842] (YP_002454823.1)	98 in 96 aa
199	330	+	hypothetical protein	hypothetical protein bthur0007_54930 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111640.1)	97 in 330
200	59	-	hypothetical protein	hypothetical protein bthur0007_54920 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111639.1)	98 in 59 aa
201	478	+	transposase for insertion sequence element IS231B	transposase for insertion sequence element IS231B [Bacillus thuringiensis MC28] (YP_006815473.1)	95 in 478 aa
202	57	-	hypothetical protein	hypothetical protein BCG9842_0161 [Bacillus cereus G9842] (YP_002454827.1)	100 in 57 aa
203	293	-	DNA integration/recombination/ invertion protein	DNA integration/recombination/invertion protein [Bacillus cereus G9842] (YP_002454828.1)	98 in 293 aa
204	136	-	hypothetical protein	hypothetical protein bthur0001_56580 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149066.1)	99 in 136 aa
205	163	-	hypothetical protein	hypothetical protein BCG9842_0164 [Bacillus cereus G9842] (YP_002454830.1)	95 in 163 aa
206	245	-	hypothetical protein	hypothetical protein BCG9842_0165 [Bacillus cereus G9842] (YP_002454831.1)	90 in 245 aa
207	313	-	pseudogene	(DNA topoisomerase III, 146kbp fragment insertion site)	

208	106	-	pseudogene		
209	324	-	transmembrane anti-sigma factor	hypothetical protein BMQ_pBM50024 [Bacillus megaterium QM B1551] (YP_003566043.1)	73 in 325 aa
210	176	-	RNA polymerase sigma factor	RNA polymerase sigma factor [Bacillus cereus Rock3-28] (ZP_04236331.1)	84 in 175 aa
211	199	-	ABC transporter permease protein	ABC transporter permease protein [Bacillus thuringiensis IBL 200] (ZP_04075478.1)	74 in 198 aa
212	621	-	ABC transporter permease protein	hypothetical protein IE9_05147 [Bacillus cereus BAG4X12-1] (ZP_17415947.1)	71 in 621 aa
213	317	-	ABC transporter ATP-binding protein	hypothetical protein ICE_05420 [Bacillus cereus BAG1X1-2] (ZP_17364930.1)	72 in 227
214	100	-	hypothetical protein	hypothetical protein IC1_06304 [Bacillus cereus VD022] (ZP_17341827.1)	49 in 97 aa
215	111	-	hypothetical protein	hypothetical protein bthur0011_5800 [Bacillus thuringiensis serovar huazhongensis BGSC 4BD1] (ZP_04082919.1)	44 in 107 aa
216	480	-	Calcineurin-like phosphoesterase	hemagglutinin-related protein [Bacillus thuringiensis serovar israelensis] (YP_001573868.1)	44 in 490 aa
217	85	-	hypothetical protein	hypothetical protein MC28_D172 [Bacillus thuringiensis MC28] (YP_006815460.1)	84 in 85 aa
218	98	-	pseudogene		
219	194	+	hypothetical protein	hypothetical protein bthur0003_63770 [Bacillus thuringiensis serovar thuringiensis str. T01001] (ZP_04137140.1)	81 in 194 aa
220	101	+	ArsR family transcriptional regulator	hypothetical protein IKM_06008 [Bacillus cereus VDM022] (ZP_17641027.1)	85 in 101 aa
221	93	+	AbrB family transcriptional regulator	AbrB family transcriptional regulator [Bacillus cereus VD045] (ZP_17566700.1)	88 in 93 aa
222	62	+	RNA chaperone Hfq	HfQ protein (RNA-binding protein) [Bacillus cereus Q1] (YP_002533362.1)	90 in 62 aa
223	65	+	hypothetical protein	hypothetical protein YBT020_27704 [Bacillus thuringiensis serovar finitimus YBT-020] (YP_005569122.1)	91 in 65 aa
224	197	-	camelysin	camelysin [Bacillus thuringiensis serovar thuringiensis str. T01001] (ZP_04137145.1)	96 in 197 aa

225	112	-	hypothetical protein	hypothetical protein bthur0007_9040 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04107101.1)	89 in 101 aa
226	109	-	hypothetical protein	hypothetical protein IKM_06002 [Bacillus cereus VDM022] (ZP_17641021.1)	39 in 99 aa
227	241	-	hypothetical protein	hypothetical protein BFZC1_00135 [Lysinibacillus fusiformis ZC1] (ZP_07047732.1)	41 in 189 aa
228	549	-	Phage integrase family protein	Phage integrase family protein [Lysinibacillus fusiformis ZC1] (ZP_07047733.1)	33 in 564 aa
229	231	-	cell envelope-bound metalloprotease	hypothetical protein bcere0013_32640 [Bacillus cereus BDRD-ST26] (ZP_04268721.1)	28 in 428 aa
230	413	-	phage integrase family site-specific recombinase	hypothetical protein IIK_01395 [Bacillus cereus VD102] (ZP_17580707.1)	32 in 640 aa
231	779	-	peptidase S8 and S53 subtilisin kexin sedolisin	hypothetical protein ACD_77C00511G0002 [uncultured bacterium] (EKD30588.1)	48 in 316 aa
232	383	-	ATPase AAA	ATPase [Clostridium botulinum A str. ATCC 3502] (YP_001254814.1)	94 in 673 aa
233	673	-	DNA topoisomerase III	DNA topoisomerase III [Bacillus cereus HuA2-1] (ZP_17486349.1)	78 in 76 aa
234	236	-	pseudogene		
235	54	-	pseudogene		
236	261	+	pseudogene		
237	131	-	HTH domain-containing DNA-binding protein	hypothetical protein bthur0006_5990 [Bacillus thuringiensis serovar kurstaki str. T03a001] (ZP_04113288.1)	87 in 105 aa
238	684	+	cry19Bb1	Pesticidal crystal protein cry19Ba (O86170.1)	74 in 690 aa
239	172	+	IS3-family transposase OrfB	transposase, orfA ISRSO11-related [Bacillus cereus E33L] (YP_245796.1)	100 in 172 aa
240	265	+	ISPsy9 transposase OrfA	transposase orfB, IS150-related protein [Bacillus thuringiensis HD-771] (YP_006602580.1)	100 in 265 aa
241	168	-	pseudogene		

242	120	-	pseudogene		
243	96	-	pseudogene		
244	306	-	pseudogene		
245	79	-	Phage protein	hypothetical protein MC28_1517 [Bacillus thuringiensis MC28] (YP_006828338.1)	59 in 104 aa
246	71	-	Phage protein		
247	324	-	pseudogene		
248	50	-	Phage protein	hypothetical protein II3_05736 [Bacillus cereus MC67] (ZP_17536834.1)	66 in 50 aa
249	351	+	pseudogene		
250	65	+	hypothetical protein	hypothetical protein IIO_04055 [Bacillus cereus VD115] (ZP_17594563.1)	89 in 63 aa
251	80	+	pseudogene		
252	269	-	IS3-family transposase OrfB	hypothetical protein IKA_00265 [Bacillus cereus VD169] (ZP_17622048.1)	96 in 269 aa
253	176	-	ISPsy9 transposase OrfA	hypothetical protein II3_00131 [Bacillus cereus MC67] (ZP_17531229.1)	97 in 172 aa
254	105	-	pseudogene		
255	259	-	pseudogene		
256	89	-	pseudogene		
257	26	-	pseudogene		
258	602	-	RNA-directed DNA polymerase	RNA-directed DNA polymerase [Enterococcus faecalis CH188] (ZP_05585309.1)	57 in 602
259	59	-	Holin	Holin [Bacillus thuringiensis serovar kurstaki str. T03a001] (ZP_04118140.1)	60 in 88 aa
260	74	-	Cof-like hydrolase	hypothetical protein II5_02389 [Bacillus cereus MSX-A1] (ZP_17539261.1)	96 in 74 aa
261	497	-	Cry40orf2	Cry40-like protein [Bacillus thuringiensis MC28] (YP_006815592.1）	74 in 498 aa
262	669	-	Cry73Aa	crystal protein [Bacillus thuringiensis serovar vazensis] (AFM37573.1)	86 in 669 aa
263	159	+	pseudogene		
264	334	+	pseudogene		

265	45	-	hypothetical protein	hypothetical protein bthur0014_3770 [Bacillus thuringiensis IBL 4222] (ZP_04063421.1)	72 in 46 aa
266	45	-	hypothetical protein	hypothetical protein [Bacillus anthracis str. H9401] (YP_006207480.1)	80 in 45 aa
267	100	-	pseudogene		
268	107	+	ISPsy9, transposase OrfA	hypothetical protein IKA_05207 [Bacillus cereus VD169] (ZP_17626990.1)	100 in 100 aa
269	279	+	ISPsy9, transposase OrfB	hypothetical protein IKA_05206 [Bacillus cereus VD169] (ZP_17626989.1)	99 in 279 aa
270	54	-	Phage protein	Phage protein [Bacillus sp. GeD10] (CCW09303.1)	80 in 54 aa
271	71	-	Phage protein	Phage protein [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00742080.1)	73 in 71 aa
272	138	-	hypothetical protein	Hypothetical cytosolic protein [Bacillus thuringiensis IBL 4222] (ZP_04063412.1)	95 in 138 aa
273	133	-	pseudogene		
274	171	-	pseudogene		
275	41	-	pseudogene		
276	118	-	pseudogene		
277	62	-	pseudogene		
278	104	-	hypothetical protein	hypothetical protein IGO_05580 [Bacillus cereus HuB5-5] (ZP_17525503.1)	55 in 94 aa
279	127	-	hypothetical protein	hypothetical protein ICU_04017 [Bacillus cereus BAG2X1-1] (ZP_17375524.1)	41 in 122
280	226	-	IS231-like transposase	Transposase for insertion sequence element IS231B [Bacillus mycoides Rock1-4] (ZP_04166736.1)	91 in 226
281	347	+	pseudogene		
282	93	-	pseudogene		
283	93	-	pseudogene		
284	119	-	copper amine oxidase-like domain-containing protein	hypothetical protein IIM_05100 [Bacillus cereus VD107] (ZP_17590246.1)	77 in 119 aa

285	84	-	transporter	Transporter [Bacillus mycoides Rock3-17] (ZP_04155868.1)	84 in 83 aa
286	45	-	hypothetical protein	hypothetical protein IEI_02611 [Bacillus cereus BAG5X2-1] (ZP_17436268.1)	70 in 56
287	340	-	mosquitocidal toxin gene	hypothetical protein bthur0009_54170 [Bacillus thuringiensis serovar andalousiensis BGSC 4AW1] (ZP_04099749.1)	32 in 341
288	478	+	IS231-like transposase	transposase for insertion sequence element IS231F, partial [Bacillus cereus VDM022] (ZP_17640768.1)	91 in 478 aa
289	226	-	IS231-like transposase	Transposase for insertion sequence element IS231B [Bacillus mycoides Rock1-4] (ZP_04166736.1)	91 in 226
290	293	-	transposase	Transposase, IS204/IS1001/IS1096/IS1165 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149049.1)	84 in 120 aa
291	127	-	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89 aa
292	144	-	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1|)	45 in 141 aa
293	708	-	Penicillin-binding protein transpeptidase	hypothetical protein IEE_02638 [Bacillus cereus BAG5X1-1] (ZP_17430747.1)	85 in 707 aa
294	235	-	peptidase M15B and M15C DD-carboxypeptidase VanY/endolysin	hypothetical protein ICG_05453 [Bacillus cereus BAG1X1-3] (ZP_17370831.1)	80 in 235
295	298	-	Serine-type D-Ala-D-Ala carboxypeptidase	hypothetical protein ICG_05452 [Bacillus cereus BAG1X1-3] (ZP_17370830.1)	78 in 298
296	384	-	sensor histidine kinase VanS	hypothetical protein ICG_05451 [Bacillus cereus BAG1X1-3] (ZP_17370829.1)	77 in 384
297	235	-	two-component response regulator VanR	hypothetical protein IC3_04873 [Bacillus cereus VD142] (ZP_17347204.1)	83 in 235
298	171	-	Invasion protein IagB domain protein	hypothetical protein IEQ_02103 [Bacillus cereus BAG6X1-2] (ZP_17459015.1)	81 in 171
299	278	-	Peptidoglycan N-acetylglucosamine deacetylase	hypothetical protein ICG_05448 [Bacillus cereus BAG1X1-3] (ZP_17370826.1)	82 in 278
300	57	-	hypothetical protein	hypothetical protein IIM_01620 [Bacillus cereus VD107] (ZP_17586766.1)	75 in 51
301	208	-	IS3 family transposase orfB	Integrase [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149206.1)	87 in 208

302	300	-	ISL3 family transposase	hypothetical protein [Bacillus thuringiensis] (WP_000098666.1)	87 in 167 aa
303	182	-	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89
304	144	-	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	45 in 141
305	431	+	IS232 transposase-like protein IstA	transposase for insertion sequence element IS232 [Bacillus cereus VD154] (ZP_17603282.1)	91 in 431 aa
306	86	+	IS232 transposase-like protein IstB	Insertion sequence IS232 putative ATP-binding protein [Bacillus thuringiensis serovar kurstaki str. T03a001] (ZP_04117980.1)	97 in 86
307	378	-	Transposase IS116/IS110/IS902	hypothetical protein IIA_05308 [Bacillus cereus VD014] (ZP_17559904.1)	76 in 403
308	110	-	transposase of ISAar40, IS3 family, IS3 group, orfA	transposase of ISAar40, IS3 family, IS3 group, orfA [Bacillus thuringiensis serovar finitimus YBT-020]	88 in 78 aa
309	260	-	Mono-ADP-ribosyltransferase C3	hypothetical protein IC3_04863 [Bacillus cereus VD142] (ZP_17347194.1)	43 in 251
310	321	-	Serine-type D-Ala-D-Ala carboxypeptidase	Serine-type D-Ala-D-Ala carboxypeptidase [Bacillus pseudogenemycoides DSM 12442] (ZP_04150951.1)	59 in 329
311	333	-	amino-acid racemase	hypothetical protein ICG_05446 [Bacillus cereus BAG1X1-3] (ZP_17370824.1)	86 in 333
312	360	-	alanine racemase domain-containing protein	hypothetical protein ICG_05445 [Bacillus cereus BAG1X1-3] (ZP_17370823.1)	84 in 360 aa
313	457	-	UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase	UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase [Bacillus cereus VD107] (ZP_17586763.1)	85 in 457
314	357	-	D-alanine-D-alanine ligase	D-alanine-D-alanine ligase [Bacillus cereus BAG1X1-3] (ZP_17370821.1)	83 in 357
315	256	-	pseudogene		
316	234	-	pseudogene		
317	112	-	transposase of ISAar40, IS3 family, IS3 group, orfA	transposase of ISAar40, IS3 family, IS3 group, orfA [Bacillus thuringiensis serovar finitimus YBT-020] (YP_005563640.1)	90 in 81 aa

318	144	+	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	55 in 141 aa
319	127	+	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89 aa
320	293	+	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149049.1)	84 in 120 aa
321	255	+	pseudogene		
322	299	-	35.8-kilodalton mosquitocidal toxin	hypothetical protein bthur0007_54850 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111632.1)	51 in 300 aa
323	249	-	hypothetical protein	hypothetical protein bthur0014_51840 [Bacillus thuringiensis IBL 4222] (ZP_04068138.1)	96 in 103 aa
324	163	-	pseudogene		
325	426	-	Collagen triple helix repeat-containing protein	hypothetical protein BCK_12595 [Bacillus cereus FRI-35] (YP_006596620.1)	45 in 410
326	294	-	pseudogene		
327	101	-	transposase	hypothetical protein ICE_05215 [Bacillus cereus BAG1X1-2] (ZP_17364725.1)	99 in 101 aa
328	256	+	mosquitocidal toxin gene	hypothetical protein bthur0012_54310 [Bacillus thuringiensis serovar pulsiensis BGSC 4CC1] (ZP_04081751.1)	30 in 259 aa
329	207	-	pseudogene		
330	403	-	Transposase IS116/IS110/IS902	hypothetical protein IIA_05308 [Bacillus cereus VD014] (ZP_17559904.1)	81 in 403 aa
331	66	-	pseudogene		
332	95	-	Phage protein	hypothetical protein bthur0014_3830 [Bacillus thuringiensis IBL 4222] (ZP_04063427.1)	87 in 95 aa
333	81	-	Phage protein	hypothetical protein II5_04377 [Bacillus cereus MSX-A1] (ZP_17541249.1)	94 in 81 aa
334	175	-	pseudogene		

335	79	-	pseudogene		
336	324	+	pseudogene		
337	93	-	hypothetical protein	hypothetical protein II5_04381 [Bacillus cereus MSX-A1] (ZP_17541253.10)	88 in 93 aa
338	95	-	pseudogene		
339	385	-	pseudogene		
340	190	+	cryBP1 family protein	hypothetical protein MC28_E061 [Bacillus thuringiensis MC28] (YP_006815559.1)	63 in 152 aa
341	411	-	pseudogene		
342	75	-	hypothetical protein	hypothetical protein MC28_E058 [Bacillus thuringiensis MC28] (YP_006815556.1)	79 in 75 aa
343	97	-	hypothetical protein	hypothetical protein MC28_E062 [Bacillus thuringiensis MC28] (YP_006815560.1)	96 in 68 aa
344	424	-	Retron-type reverse transcriptase	hypothetical protein bthur0005_3190 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04118566.1)	95 in 424 aa
345	431	+	IS232 family transposase IstA	transposase for insertion sequence element IS232 [Bacillus cereus VD154] (ZP_17603282.1)	91 in 431 aa
346	250	+	IS232 family transposase IstB	hypothetical protein IK5_00386 [Bacillus cereus VD154] (ZP_17603283.1)	98 in 250 aa
347	157	+	pseudogene		
348	722	+	Cry20Bb1	Cry20-like delta endotoxin [Bacillus thuringiensis] (ACS93601.1)	73 in 766 aa
349	187	-	pseudogene		
350	272	-	pseudogene		
351	201	-	pseudogene		
352	375	-	spore germination protein GerZC	putative spore germination protein [Bacillus thuringiensis serovar israelensis] (YP_001573812.1)	85 in 375 aa
353	367	-	spore germination protein GerZB	putative spore germination protein [Bacillus thuringiensis serovar israelensis] (YP_001573813.1)	90 in 367 aa

354	491	-	spore germination protein GerZA	putative spore germination protein [Bacillus thuringiensis serovar israelensis] (YP_001573814.1)	95 in 490 aa
355	190	-	pseudogene		
356	194	-	Cry20-like delta endotoxin	hypothetical protein MC28_D170 [Bacillus thuringiensis MC28] (YP_006815458.1)	50 in 212 aa
357	456	-	Calcineurin-like phosphoesterase	hemagglutinin-related protein [Bacillus thuringiensis serovar israelensis] (YP_001573868.1)	52 in 456
358	85	-	hypothetical protein	hypothetical protein MC28_D172 [Bacillus thuringiensis MC28] (YP_006815460.1)	82 in 85 aa
359	128	-	pseudogene		
360	253	-	pseudogene	(DNA topoisomerase III, 146kbp fragment insertion site)	
361	121	-	DNA topoisomerase TopB	DNA topoisomerase 3 [Bacillus cereus G9842] (YP_002454834.1)	98 in 121 aa
362	60	-	pseudogene		
363	304	-	ATPase AAA	stage V sporulation protein K [Bacillus cereus G9842] (YP_002454835.1)	99 in 304 aa
364	241	-	ATPase AAA+	stage V sporulation protein K [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111626.1)	95 in 239
365	140	-	hypothetical protein	hypothetical protein bthur0004_60870 [Bacillus thuringiensis serovar sotto str. T04001] (ZP_04130236.1)	96 in 140 aa
366		-	hypothetical protein	hypothetical protein BCG9842_0172 [Bacillus cereus G9842] (YP_002454837.1)	97 in 338 aa
367	458	+	nlpC/P60 family protein	hypothetical protein IK9_05470 [Bacillus cereus VD166] (ZP_17621143.1)	98 in 458 aa
368	216	-	hypothetical protein	hypothetical protein bthur0001_53860 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04148816.1)	95 in 216 aa
369	1048	-	cell wall endopeptidase, family M23/M37	cell wall endopeptidase, family M23/M37 [Bacillus cereus G9842] (YP_002454840.1)	99 in 1048 aa
370	1562	-	Reticulocyte binding protein	hypothetical protein BCG9842_0176 [Bacillus cereus G9842] (YP_002454841.1)	99 in 1562 aa

371	698	-	hypothetical protein	hypothetical protein bthur0007_54690 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111616.1)	99 in 698 aa
372	301	-	hypothetical protein	hypothetical protein IK9_05465 [Bacillus cereus VD166] (ZP_17621138.1)	99 in 301 aa
373	108	-	hypothetical protein	hypothetical protein BCG9842_0179 [Bacillus cereus G9842] (YP_002454844.1)	97 in 108 aa
374	216	-	hypothetical protein	hypothetical protein bthur0007_54660 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111613.1)	99 in 216 aa
375	142	-	hypothetical protein	hypothetical protein BCG9842_0181 [Bacillus cereus G9842] (YP_002454846.1)	99 in 142 aa
376	96	+	ATP synthase F0 subunit 6	hypothetical protein BCG9842_0182 [Bacillus cereus G9842] (YP_002454847.1)	98 in 96 aa
377	433	-	DNA primase	putative DNA primase [Bacillus cereus G9842] (YP_002454848.1)	98 in 433 aa
378	551	-	hypothetical protein	hypothetical protein BCG9842_0184 [Bacillus cereus G9842] (YP_002454849.1)	97 in 551 aa
379	85	-	hypothetical protein	hypothetical protein BCG9842_0185 [Bacillus cereus G9842] (YP_002454850.1)	92 in 73 aa
380	379	+	DNA polymerase III subunit beta	DNA polymerase III, beta subunit [Bacillus cereus G9842] (YP_002454851.1)	99 in 379 aa
381	137	+	hypothetical protein	hypothetical protein BCG9842_0187 [Bacillus cereus G9842] (YP_002454852.1)	99 in 137 aa
382	178	+	hypothetical protein	hypothetical protein BCG9842_0188 [Bacillus cereus G9842] (YP_002454853.1)	99 in 178 aa
383	227	+	hypothetical protein	hypothetical protein BCG9842_0189 [Bacillus cereus G9842] (YP_002454854.1)	97 in 227 aa
384	406	+	hypothetical protein	hypothetical protein BCG9842_0190 [Bacillus cereus G9842] (YP_002454855.1)	96 in 406 aa
385	307	+	ThiF family protein	ThiF family protein [Bacillus cereus G9842] (YP_002454856.1)	99 in 307 aa
386	337	-	hypothetical protein	hypothetical protein BCG9842_0192 [Bacillus cereus G9842] (YP_002454857.1)	98 in 337 aa
387	234	-	hypothetical protein	hypothetical protein BCG9842_0193 [Bacillus cereus G9842] (YP_002454858.1)	98 in 222 aa
388	136	-	hypothetical protein	hypothetical protein BCG9842_0194 [Bacillus cereus G9842] (YP_002454859.1)	100 in 136 aa
389	83	-	hypothetical protein	hypothetical protein BCG9842_0195 [Bacillus cereus G9842] (YP_002454860.1)	100 in 83 aa
390	81	-	hypothetical protein	hypothetical protein bthur0007_54480 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111595.1)	100 in 81 aa
391	81	-	hypothetical protein	hypothetical protein BCG9842_0197 [Bacillus cereus G9842] (YP_002454862.1)	100 in 81 aa

392	308	-	hypothetical protein	hypothetical protein bthur0007_54460 [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04111593.1)	97 in 308 aa
393	327	-	hypothetical protein	conserved membrane protein, putative [Bacillus cereus G9842] (YP_002454864.1)	98 in 327 aa
394	498	-	bacterial type II/IV secretion system protein	bacterial type II/IV secretion system protein [Bacillus cereus G9842] (YP_002454865.1)	97 in 498 aa
395	272	-	SAF domain family protein	hypothetical protein BCG9842_0201 [Bacillus cereus G9842] (YP_002454866.1)	98 in 272 aa
396	303	-	flp pilus assembly protein CpaB	hypothetical secreted protein [Bacillus cereus G9842] (YP_002454867.1)	99 in 303 aa
397	159	-	hypothetical protein	hypothetical protein BCG9842_0203 [Bacillus cereus G9842] (YP_002454868.1)	100 in 159 aa
398	863	-	PQQ enzyme repeat domain protein	hypothetical protein BCG9842_0204 [Bacillus cereus G9842] (YP_002454869.1)	99 in 863 aa
399	242	-	hypothetical protein	hypothetical protein BCG9842_0205 [Bacillus cereus G9842] (YP_002454870.1)	96 in 248 aa
400	45	-	hypothetical protein	hypothetical protein BCG9842_0206 [Bacillus cereus G9842] (YP_002454871.1)	91 in 45 aa
401	214	-	hypothetical protein	hypothetical protein BCG9842_0207 [Bacillus cereus G9842] (YP_002454872.1)	95 in 213 aa
402	182	-	hypothetical protein	conserved hypothetical membrane protein, putative [Bacillus cereus G9842] (YP_002454873.1)	100 in 182 aa
403	507	-	FtsZ/tubulin-related protein	putative FtsZ/tubulin-related protein [Bacillus cereus G9842] (YP_002454874.1)	99 in 507 aa
404	142	-	hypothetical protein	hypothetical protein BCG9842_0210 [Bacillus cereus G9842] (YP_002454875.1)	99 in 142 aa
405	51	-	hypothetical protein	ypothetical protein BCG9842_0211 [Bacillus cereus G9842] (YP_002454876.1)	94 in 48 aa
406	509	+	hypothetical protein	hypothetical protein BCG9842_0213 [Bacillus cereus G9842] (YP_002454878.1)	99 in 509 aa
407	962	+	TraG/TraD family conjugation protein	hypothetical protein IK9_05431 [Bacillus cereus VD166] (ZP_17621104.1)	94 in 962 aa
408	447	-	replicative DNA helicase (ori-related genes)	replicative DNA helicase [Bacillus cereus G9842] (YP_002454881.1)	99 in 447 aa
409	178	+	RsfA family transcription factor (ori-related genes)	transcription factor, RsfA family [Bacillus cereus G9842] (YP_002454882.1)	96 in 178 aa

410	242	-	MerR family transcriptional regulator (ori-related genes)	hypothetical protein BCG9842_0218 [Bacillus cereus G9842] (YP_002454883.1)	99 in 242 aa

Table S2. The CDS in pMOGI222 and their annotations.

CDS	size (aa)	Strand	annotation	Best hit in databases (GenBank no.)	(% aa identity)
1	129	-	hypothetical protein (ori-related genes)	hypothetical protein IIM_05167 [Bacillus cereus VD107] (ZP_17590313.1)	80 in 129 aa
2	275	-	chromosome partitioning ATPase (ori-related genes)	ATPase [Bacillus cereus] (WP_000335378.1)	90 in 269 aa
3	518	+	replication initiation protein (ori-related genes)	replication initiation protein [Bacillus cereus] (WP_001099049.1)	97 in 518 aa
4	88	+	pseudogene		
5	710	-	Tn3 family transposase	hypothetical protein IIO_06154 [Bacillus cereus VD115] (ZP_17596662.1)	75 in 667 aa
6	284	+	TnpI resolvase	TnpI resolvase [Bacillus thuringiensis] (YP_001485222.1)	92 in 284 aa
7	987	+	transposase for Tn1546	transposase [Bacillus cereus] (YP_001966662.1)	98 in 987 aa
8	478	+	transposase for insertion sequence element IS231B	transposase for insertion sequence element IS231B [Bacillus thuringiensis MC28] (YP_006815473.1)	95 in 478 aa
9	169	+	pseudogene		
10	215	-	DNA integration/recombination/ inversion protein	DNA integration/recombination/inversion protein [Bacillus thuringiensis IBL 4222] (ZP_04069695.1)	96 in 209 aa
11	74	+	hypothetical protein	hypothetical protein IKG_05542 [Bacillus cereus VD200] (ZP_17633897.1)	81 in 75 aa
12	118	+	hypothetical protein	hypothetical protein pBt066 [Bacillus thuringiensis serovar israelensis] (YP_001573800.1)	94 in 118 aa
13	49	+	hypothetical protein	hypothetical protein IKM_05534 [Bacillus cereus VDM022] (ZP_17640732.1)	78 in 49 aa
14	305	-	chromosome segregation ATPase	hypothetical protein bthur0005_53640 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123497.1)	90 in 309 aa

15	102	-	hypothetical protein	hypothetical protein MC28_C058 [Bacillus thuringiensis MC28] (YP_006815245.1)	95 in 102 aa
16	96	-	ArsR family transcriptional regulator	hypothetical protein IC1_05970 [Bacillus cereus VD022] (ZP_17341493.1)	82 in 94 aa
17	92	+	small DNA-binding protein	small DNA-binding protein [Bacillus thuringiensis serovar israelensis] (YP_001573818.1)	91 in 92 aa
18	57	-	RNA chaperone Hfq	RNA chaperone Hfq [Bacillus thuringiensis IBL 4222] (ZP_04069280.1)	98 in 52 aa
19	93	-	AbrB family transcriptional regulator	putative transcriptional regulator [Bacillus cereus Q1] (YP_002533363.1)	84 in 93 aa
20	61	+	hypothetical protein	hypothetical protein pBt095 [Bacillus thuringiensis serovar israelensis] (YP_001573821.1)	80 in 61 aa
21	260	-	CAAX amino terminal protease family protein	hypothetical protein II3_05177 [Bacillus cereus MC67] (ZP_17536275.1)	37 in 240 aa
22	710	-	DNA topoisomerase III	DNA topoisomerase III [Bacillus cereus VD154] (ZP_17609214.1)	96 in 445 aa
23	219	-	hypothetical protein	hypothetical protein bthur0005_53750 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123508.1)	98 in 219 aa
24	102	-	hypothetical protein	hypothetical protein bthur0005_53760 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123509.1)	100 in 102 aa
25	269	-	hypothetical protein	hypothetical protein bthur0005_53770 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123510.1)	99 in 269 aa
26	53	-	hypothetical protein	hypothetical protein IK5_06321, partial [Bacillus cereus VD154] (ZP_17609218.1)	100 in 53 aa
27	81	-	hypothetical protein	hypothetical protein MC28_D003 [Bacillus thuringiensis MC28] (YP_006815291.1)	89 in 81 aa
28	178	-	hypothetical protein	hypothetical protein IK5_05725 [Bacillus cereus VD154] (ZP_17608622.1)	97 in 178 aa
29	416	-	hypothetical protein	hypothetical protein IK5_05727 [Bacillus cereus VD154] (ZP_17608624.1)	96 in 416 aa

30	195	-	hypothetical protein	hypothetical protein IK5_05728 [Bacillus cereus VD154] (ZP_17608625.1)	94 in 195 aa
31	342	-	conjugation protein TraL	hypothetical protein IK5_05729 [Bacillus cereus VD154] (ZP_17608626.1)	99 in 342 aa
32	304	-	hypothetical protein	hypothetical protein bthur0005_60130 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04124026.1)	100 in 304 aa
33	301	-	hypothetical protein	hypothetical protein bthur0005_60120 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04124025.1)	98 in 301 aa
34	765	-	conjugation protein TrsK	hypothetical protein IK5_05732, partial [Bacillus cereus VD154] (ZP_17608629.1)	95 in 393 aa
35	605	-	conjugal transfer protein TraE	protein TrsE [Bacillus sp. GeD10] (CCW06843.1)	99 in 605 aa
36	224	-	hypothetical protein	hypothetical protein IK5_05737 [Bacillus cereus VD154] (ZP_17608634.1)	97 in 224 aa
37	106	-	hypothetical protein	hypothetical protein IK5_05738 [Bacillus cereus VD154] (ZP_17608635.1)	100 in 106 aa
38	117	-	hypothetical protein	hypothetical protein IK5_05739 [Bacillus cereus VD154] (ZP_17608636.1)	100 in 117 aa
39	58	-	hypothetical protein	hypothetical protein IK5_05740 [Bacillus cereus VD154] (ZP_17608637.1)	97 in 58 aa
40	504	-	Surface adhesion protein; Bacillolysin / Insecticidal delta-endotoxin protein	hypothetical protein IK5_05741 [Bacillus cereus VD154] (ZP_17608638.1)	96 in 505 aa
41	242	-	hypothetical protein	hypothetical protein bthur0005_57990 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123853.1)	97 in 242 aa
42	81	-	hypothetical protein	hypothetical protein EBGED10_35650 [Bacillus sp. GeD10] (CCW06835.1)	93 in 81 aa
43	154	-	TRAG family protein	hypothetical protein bthur0005_57970 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123851.1)	88 in 154 aa
44	153	-	hypothetical protein	hypothetical protein bthur0005_58140 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123864.1)	92 in 153 aa
45	751	-	TQXA domain-containing protein	TQXA domain-containing protein [Bacillus cereus VDM062] (ZP_17653212.1)	64 in 747 aa

46	69	-	Xre family transcriptional regulator	Transcriptional regulator, Xre [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123550.1)	96 in 69 aa
47	119	+	MerR family transcriptional regulator	hypothetical protein IK5_06346 [Bacillus cereus VD154] (ZP_17609243.1)	100 in 119 aa
48	409	+	hypothetical protein	hypothetical protein IK5_05988 [Bacillus cereus VD154] (ZP_17608885.1)	87 in 410 aa
49	166	+	hypothetical protein	hypothetical protein III_06028 [Bacillus cereus VD078] (ZP_17579226.1)	77 in 176 aa
50	408	+	hypothetical protein	hypothetical protein III_05946 [Bacillus cereus VD078] (ZP_17579144.1)	98 in 408 aa
51	149	+	hypothetical protein	hypothetical protein IK5_06231 [Bacillus cereus VD154] (ZP_17609128.1)	91 in 149 aa
52	71	+	hypothetical protein	hypothetical protein bthur0005_51970 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123363.1)	90 in 60 aa
53	80	-	hypothetical protein	hypothetical protein bthur0005_52320 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123394.1)	93 in 43 aa
54	98	-	hypothetical protein	hypothetical protein IK5_05757 [Bacillus cereus VD154] (ZP_17608654.1)	96 in 98 aa
55	109	-	hypothetical protein	hypothetical protein IK5_05758 [Bacillus cereus VD154] (ZP_17608655.1)	98 in 109 aa
56	86	-	hypothetical protein	hypothetical protein IK5_05759 [Bacillus cereus VD154] (ZP_17608656.1)	98 in 86 aa
57	949	+	MobA/MobL family protein	hypothetical protein IK5_05760 [Bacillus cereus VD154] (ZP_17608657.1)	98 in 949 aa
58	216	+	hypothetical protein	hypothetical protein bthur0005_60700 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04124077.1)	96 in 216 aa
59	48	-	hypothetical protein	hypothetical protein IK5_05762 [Bacillus cereus VD154] (ZP_17608659.1)	100 in 48 aa
60	348	-	response regulator aspartate phosphatase I	Prophage LambdaBa01, TPR domain protein [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04124075.1)	98 in 348 aa
61	98	+	hypothetical protein	hypothetical protein bthur0005_60670 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04124074.1)	98 in 98 aa
62	61	+	hypothetical protein	hypothetical protein IK5_06005 [Bacillus cereus VD154] (ZP_17608902.1)	97 in 61 aa
63	393	+	FtsK/SpoIIIE ATPase	hypothetical protein IK5_05766 [Bacillus cereus VD154] (ZP_17608663.1)	94 in 393 aa

64	205	+	Phage protein	hypothetical protein IK5_05767 [Bacillus cereus VD154] (ZP_17608664.1)	94 in 204 aa
65	75	-	XRE family transcriptional regulator	conserved hypothetical protein [Geobacillus sp. G11MC16] (ZP_03149500.1)	65 in 74 aa
66	72	-	hypothetical protein	hypothetical protein IK5_06243 [Bacillus cereus VD154] (ZP_17609140.1)	81 in 73 aa
67	56	-	hypothetical protein	hypothetical protein EBGED10_14800 [Bacillus sp. GeD10] (CCW04762.1)	73 in 56 aa
68	79	-	hypothetical protein	hypothetical protein IK5_06245 [Bacillus cereus VD154] (ZP_17609142.1)	85 in 79 aa
69	78	-	hypothetical protein	hypothetical protein bmyco0002_58430 [Bacillus mycoides Rock1-4] (ZP_04166458.1)	42 in 79 aa
70	82	-	hypothetical protein	hypothetical protein III_05964 [Bacillus cereus VD078] (ZP_17579162.1)	75 in 81 aa
71	183	-	hypothetical protein	hypothetical protein bthur0005_55920 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123674.1)	89 in 183 aa
72	157	+	hypothetical protein	hypothetical protein [Bacillus cereus](WP_016099623.1)	99 in 157 aa
73	414	-	tetratricopeptide repeat family protein	hypothetical protein bmyco0002_59880 [Bacillus mycoides Rock1-4] (ZP_04166587.1)	89 in 404 aa
74	130	-	TIR_2 superfamily protein	hypothetical protein bmyco0002_59870 [Bacillus mycoides Rock1-4] (ZP_04166586.1)	94 in 130 aa
75	61	-	pseudogene		
76	234	+	hypothetical protein	conserved hypothetical protein [Bacillus cereus 03BB108] (ZP_03112632.1)	92 in 234 aa
77	126	-	hypothetical protein	hypothetical protein RBTH_07022 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00739166.1)	74 in 125 aa
78	68	-	Cro/CI family transcriptional regulator	Transcriptional regulator, Cro/CI family [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00739165.1)	85 in 68 aa
79	298	+	Secreted subtilisin-like serine protease	protease [Bacillus cereus 03BB108] (ZP_03112649.1)	80 in 286 aa
80	96	+	Thiol-disulfide oxidoreductase BdbC	Thiol-disulfide oxidoreductase BdbC [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00739163.1)	84 in 96 aa

81	153	+	Thioredoxin	Thioredoxin [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00739162.1)	84 in 153 aa
82	71	-	hypothetical protein	hypothetical protein IK5_06243 [Bacillus cereus VD154] (ZP_17609140.1)	63 in 73 aa
83	171	-	hypothetical protein	hypothetical protein MC28_E153 [Bacillus thuringiensis MC28] (YP_006815651.1)	85 in 132 aa
84	99	-	hypothetical protein	hypothetical protein MC28_E152 [Bacillus thuringiensis MC28] (YP_006815650.1)	95 in 99 aa
85	49	+	hypothetical protein	hypothetical protein MC28_E151 [Bacillus thuringiensis MC28] (YP_006815649.1)	96 in 49 aa
86	118	+	hypothetical protein	hypothetical protein MC28_E150 [Bacillus thuringiensis MC28] (YP_006815648.1)	90 in 120 aa
87	148	-	hypothetical protein	hypothetical protein MC28_E149 [Bacillus thuringiensis MC28] (YP_006815647.1)	73 in 158 aa
88	190	+	hypothetical protein	hypothetical protein MC28_E147 [Bacillus thuringiensis MC28](YP_006815645.1)	81 in 190 aa
89	56	+	pseudogene		
90	144	+	hypothetical protein	hypothetical protein BTF1_31851 [Bacillus thuringiensis HD-789] (YP_006613948.1)	100 in 144 aa
91	131	+	hypothetical protein	hypothetical protein RBTH_07781 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00738434.1)	100 in 131 aa
92	137	+	Cobalamin synthesis protein P47K	hypothetical protein RBTH_07780 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00738433.1)	85 in 137 aa
93	390	+	lipase	lipase [Bacillus thermoamylovorans] (BAH70300.1)	57 in 375 aa
94	284	+	TnpI resolvase	TnpI resolvase [Bacillus thuringiensis] (YP_001485222.1)	92 in 284 aa

95	987	+	transposase for transposon Tn1546	transposase [Bacillus cereus] (YP_001966662.1)	98 in 987 aa
96	53	+	hypothetical protein	hypothetical protein bthur0013_55230 [Bacillus thuringiensis IBL 200] (ZP_04075187.1)	72 in 53 aa
97	260	+	hypothetical protein	hypothetical protein BSSC8_22580 [Bacillus subtilis subsp. subtilis str. SC-8] (ZP_12671314.1)	32 in 230 aa
98	111	+	hypothetical protein	hypothetical protein bpmyx0001_50440 [Bacillus pseudomycoides DSM 12442] (ZP_04154217.1)	84 in 58 aa
99	172	+	chromosome segregation ATPase	hypothetical protein IK5_06171 [Bacillus cereus VD154] (ZP_17609068.1)	91 in 176 aa
100	194	-	DNA-Invertase BINR	hypothetical protein IKM_05568 [Bacillus cereus VDM022] (ZP_17640766.1)	83 in 187 aa
101	74	-	hypothetical protein	hypothetical protein SSIL_2306 [Solibacillus silvestris StLB046] (YP_006462875.1)	68 in 74 aa
102	199	-	hypothetical protein	hypothetical protein BN424_3862 [Carnobacterium maltaromaticum LMA28] (YP_006994581.1)	67 in 183 aa
103	313	-	epoxide hydrolase 2	epoxide hydrolase 2 [Microscilla marina ATCC 23134] (ZP_01690851.1)	83 in 307 aa
104	309	-	DeoR family transcriptional regulator	hypothetical protein IYC_05053 [Clostridium sporogenes PA 3679] (ZP_18252214.1)	47 in 310 aa
105	183	+	Resolvase	Resolvase [Bacillus thuringiensis IBL 200] (ZP_04075383.1)	97 in 182 aa
106	220	-	mosquitocidal toxin	RecName: Full=Pesticidal crystal protein cry19Ba; AltName: Full=78 kDa crystal protein; AltName: Full=Crystaline entomocidal protoxin; AltName: Full=Insecticidal delta-endotoxin CryXIXB(a) (O86170.1)	97 in 217 aa
107	139	+	pseudogene		
108	825	+	Cry27Ab1	RecName: Full=Pesticidal crystal protein cry27Aa; AltName: Full=94 kDa crystal protein; AltName: Full=Crystaline entomocidal protoxin; AltName: Full=Insecticidal delta-endotoxin CryXXVIIA(a) (Q9S597.1)	83 in 829 aa

109	160	+	group-specific protein	hypothetical protein IE5_00813 [Bacillus cereus BAG3X2-2] (ZP_17400155.1)	79 in 140 aa
110	261	+	response regulator aspartate phosphatase K	response regulator aspartate phosphatase [Bacillus thuringiensis serovar thuringiensis str. IS5056] (YP_007491988.1)	75 in 238 aa
111	478	-	transposase for insertion sequence element IS231B	transposase for insertion sequence element IS231B [Bacillus thuringiensis MC28] (YP_006815473.1)	95 in478 aa
112	260	-	hypothetical protein	hypothetical protein BSSC8_22580 [Bacillus subtilis subsp. subtilis str. SC-8] (ZP_12671314.1)	32 in 230 aa
113	53	-	hypothetical protein	hypothetical protein bthur0013_55230 [Bacillus thuringiensis IBL 200] (ZP_04075187.1)	72 in 53 aa
114	65	+	pseudogene		
115	162	-	copper amine oxidase domain-containing protein	Hypothetical protein RBTH_08755 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00741906.1)	77 in 162 aa
116	123		pseudogene		
117	460	-	Collagen-like triple helix repeat protein	Collagen-like triple helix repeat protein [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00741909.1)	60 in 359 aa
118	248	-	IS605 OrfB family transposase	transposase, OrfB family [Bacillus cereus AH1134] (ZP_03233746.1)	94 in 247 aa
119	473	-	Transposase for insertion sequence element IS231B	Transposase for insertion sequence element IS231B [Bacillus thuringiensis serovar andalousiensis BGSC 4AW1] (ZP_04099942.1)	92 in 472 aa
120	86	+	DNA binding protein	DNA binding protein [Bacillus thuringiensis serovar andalousiensis BGSC 4AW1] (ZP_04099943.1)	91 in 86 aa
121	144	+	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	45 in 141 aa
122	127	+	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89 aa

123	293	+	Transposase, IS204/IS1001/IS1096/IS1165	Transposase, IS204/IS1001/IS1096/IS1165 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149049.1)	84 in 120 aa
124	64	+	Transposase for insertion sequence element IS231B	transposase for insertion sequence element IS231E, partial [Bacillus cereus BAG6O-1] (ZP_17445153.1)	94 in 64 aa
125	115	+	Exosporium protein ExsB	hypothetical protein IIA_02388 [Bacillus cereus VD014] (ZP_17556984.1)	70 in 126 aa
126	162	-	copper amine oxidase domain-containing protein	Hypothetical protein RBTH_08755 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00741906.1)	78 in 162 aa
127	111	-	Transporter	Transporter [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00741907.1)	79 in 120 aa
128	201	-	Collagen-like triple helix repeat protein	Collagen-like triple helix repeat protein [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00741909.1)	64 in 216 aa
129	235	+	pseudogene		
130	137	-	Ice nucleation protein	putative deletion pseudogene prduct [Bacillus thuringiensis serovar israelensis] (YP_001573797.1)	61 in 135 aa
131	471	-	Calcineurin-like phosphoesterase; mosquitocidal toxin protein	hemagglutinin-related protein [Bacillus thuringiensis serovar israelensis] (YP_001573868.1)	48 in 483 aa
132	185	+	19kda accessory protein	hypothetical protein MC28_E061 [Bacillus thuringiensis MC28] (YP_006815559.1)	51 in 185 aa
133	683	+	pesticidial crystal protein cry4AA	pesticidial crystal protein cry4AA [Bacillus thuringiensis serovar finitimus YBT-020] (YP_005569289.1)	30 in 648 aa
134	506	-	crystal protein ET69	crystal protein ET69 [Bacillus thuringiensis] (Sequence ID: gb|AAK64558.1)	33 in 480 aa
135	77	-	hypothetical protein	hypothetical protein MC28_E058 [Bacillus thuringiensis MC28] (YP_006815556.1)	85 in 73 aa

136	98	-	hypothetical protein	hypothetical protein MC28_E062 [Bacillus thuringiensis MC28] (YP_006815560.1)	97 in 67 aa
137	91	-	pseudogene		
138	144	+	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	45 in 141 aa
139	127	+	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89 aa
140	293	+	pseudogene		
141	288	-	pseudogene		
142	86	-	DNA binding protein	DNA binding protein [Bacillus thuringiensis serovar andalousiensis BGSC 4AW1] (ZP_04099943.1)	91 in 86 aa
143	473	+	Transposase for insertion sequence element IS231B	Transposase for insertion sequence element IS231B [Bacillus thuringiensis serovar andalousiensis BGSC 4AW1] (ZP_04099942.1)	92 in 472 aa
144	286	+	hypothetical protein	hypothetical protein RBTH_07849 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00738488.1)	98 in 286 aa
145	352	+	hypothetical protein	hypothetical protein pBt074 [Bacillus thuringiensis serovar israelensis] (YP_001573807.1)	90 in 352 aa
146	248	+	hypothetical protein	hypothetical protein bthur0014_62240 [Bacillus thuringiensis IBL 4222] (ZP_04069141.1)	88 in 248 aa
147	92	-	DNA-binding protein HU	small DNA-binding protein [Bacillus thuringiensis serovar israelensis] (YP_001573818.1)	97 in 92 aa
148	48	+	hypothetical protein	hypothetical protein IC1_05970 [Bacillus cereus VD022] (ZP_17341493.1)	67 in 48 aa
149	89	-	transposase of ISAar40, IS3 family, IS3 group, orfA	transposase of ISAar40, IS3 family, IS3 group, orfA [Bacillus thuringiensis serovar finitimus YBT-020] (YP_005563640.1)	85 in 88 aa

150	70	-	Methionine-rich protein	hypothetical protein bpmyx0001_29000 [Bacillus pseudomycoides DSM 12442] (ZP_04152091.1)	84 in 70 aa
151	73	+	pseudogene		
152	480	-	GntR family transcriptional regulator	GntR domain-containing protein [Bacillus thuringiensis serovar israelensis] (YP_001573828.1)	98 in 480 aa
153	299	+	GHMP kinase	kinase [Bacillus thuringiensis serovar israelensis] (YP_001573827.1)	96 in 299 aa
154	235	+	alanyl-tRNA synthetase	tRNA synthetase-related protein [Bacillus thuringiensis serovar israelensis] (YP_001573826.1)	97 in 235 aa
155	345	+	pyridoxal-phosphate dependent protein	pyridoxal-phosphate dependent protein [Bacillus thuringiensis MC28] (YP_006815619.1)	96 in 345 aa
156	394	+	class-II aminotransferase	class-II aminotransferase [Bacillus thuringiensis MC28] (YP_006815618.1)	95 in 394 aa
157	299	+	EamA-like transporter family	EamA-like transporter family [Bacillus thuringiensis MC28] (YP_006815617.1)	98 in 299 aa
158	93	+	AbrB family transcriptional regulator	putative transcriptional regulator [Bacillus thuringiensis serovar israelensis] (YP_001573820.1)	91 in 89 aa
159	65	+	RNA chaperone Hfq	RNA chaperone Hfq [Bacillus thuringiensis MC28] (YP_006815242.1)	53 in 77 aa
160	83	+	transposase of ISAar40, IS3 family, IS3 group, orfA	hypothetical protein IKM_05489 [Bacillus cereus VDM022] (ZP_17640687.1)	57 in 86 aa
161	278	-	ISPsy9, transposase OrfB	hypothetical protein IKA_05206 [Bacillus cereus VD169] (ZP_17626989.1)	99 in 278 aa
162	107	-	transposase	hypothetical protein IKA_05207 [Bacillus cereus VD169] (ZP_17626990.1)	100 in 107 aa
163	101	+	transposase	hypothetical protein ICE_05215 [Bacillus cereus BAG1X1-2] (ZP_17364725.1)	99 in 101 aa
164	144	+	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	45 in 141 aa
165	127	+	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in89 aa

166	293	+	Transposase, IS204/IS1001/IS1096/IS1165	Transposase, IS204/IS1001/IS1096/IS1165 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149049.1)	84 in120 aa
167	244	+	pseudogene		
168	195	+	pseudogene		
169	190	-	antioxidant, AhpC/TSA family	antioxidant, AhpC/TSA family [Bacillus sp. M 2-6] (ZP_10162419.1)	31 in 144 aa
170	180	+	pseudogene		
171	147	-	hypothetical protein	hypothetical protein MC28_E149 [Bacillus thuringiensis MC28] (YP_006815647.1)	81 in 123 aa
172	62	+	pseudogene		
173	82	+	hypothetical protein	hypothetical protein MC28_F175 [Bacillus thuringiensis MC28] (YP_006815858.1)	96 in 82 aa
174	386	-	DNA integration/recombination/inversion protein	DNA integration/recombination/inversion protein [Bacillus thuringiensis serovar monterrey BGSC 4AJ1] (ZP_04108478.1)	95 in 386 aa
175	321	+	excisionase family DNA binding domain-containing protein	hypothetical protein IG3_06349 [Bacillus cereus HuA2-1] (ZP_17491383.1)	88 in 320 aa
176	137	-	pseudogene		
177	435	-	putative reverse transcriptase	RNA-directed DNA polymerase (Reverse transcriptase) [Bacillus thuringiensis serovar berliner ATCC 10792] (ZP_04102003.1)	99 in 435 aa
178	59	-	hypothetical protein	hypothetical protein pBMB0558_00185 [Bacillus thuringiensis CT43] (YP_004169148.1)	98 in 59 aa
179	334	-	pseudogene		
180	372	+	pseudogene		
181	59	+	hypothetical protein	hypothetical protein pBMB0558_00185 [Bacillus thuringiensis CT43] (YP_004169148.1)	98 in 59 aa

182	373	+	putative reverse transcriptase	hypothetical protein IEE_05364 [Bacillus cereus BAG5X1-1] (ZP_17433473.1)	97 in 373 aa
183	136	+	pseudogene		
184	297	+	chromosome segregation ATPase	hypothetical protein bthur0014_63540 [Bacillus thuringiensis IBL 4222] (ZP_04069259.1)	97 in 301 aa
185	116	-	hypothetical protein	hypothetical protein IKM_05468 [Bacillus cereus VDM022] (ZP_17640666.1)	99 in 116 aa
186	557	-	mosquitocidal toxin protein	hypothetical protein bthur0009_56310 [Bacillus thuringiensis serovar andalousiensis BGSC 4AW1] (ZP_04099944.1)	38 in 149 aa
187	101	+	RNA polymerase sigma factor SigX	RNA polymerase sigma factor SigX [Bacillus thuringiensis serovar israelensis] (YP_001573832.1)	93 in 72 aa
188	194	-	Transposon resolvase	hypothetical protein IKM_05559 [Bacillus cereus VDM022] (ZP_17640757.1)	99 in 194 aa
189	335	+	transposase Tn3 family protein	hypothetical protein IKM_05637, partial [Bacillus cereus VDM022] (ZP_17640835.1)	99 in 334 aa
190	431	+	transposase IstA	transposase IstA [Bacillus thuringiensis] (YP_001485232.1)	99 in 431 aa
191	250	+	ATP-binding protein IstB	ATP-binding protein IstB [Bacillus thuringiensis] (YP_001485231.1)	99 in 250 aa
192	144	+	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	45 in 141 aa
193	127	+	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89 aa
194	293	+	Transposase, IS204/IS1001/IS1096/IS1165	Transposase, IS204/IS1001/IS1096/IS1165 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149049.1)	84 in 120 aa
195	703	+	transposase Tn3 family protein	hypothetical protein IKM_05637, partial [Bacillus cereus VDM022] (ZP_17640835.1)	98 in 647 aa
196	175	+	pseudogene		
197	202	-	MerR superfamily protein	hypothetical protein IK3_05615 [Bacillus cereus VD148] (ZP_17602795.1)	78 in 466 aa

198	160	+	hypothetical protein	hypothetical protein pBT9727_0060 [Bacillus thuringiensis serovar konkukian str. 97-27] (YP_173303.1)	91 in 190 aa
199	461	-	Phosphatidylinositol-specific phospholipase	Phosphatidylinositol-specific phospholipase [Bacillus cereus AH621] (ZP_04298276.1)	78 in 466 aa
200	190	-	hypothetical protein	hypothetical protein MC28_E159 [Bacillus thuringiensis MC28] (YP_006815657.1)	91 in 190 aa
201	408	+	hypothetical protein	hypothetical protein [Bacillus cereus] (WP_016099379.1)	87 in 408 aa
202	168	-	pseudogene		
203	75	-	pseudogene		
204	293	-	Transposase, IS204/IS1001/IS1096/IS1165	Transposase, IS204/IS1001/IS1096/IS1165 [Bacillus thuringiensis serovar tochigiensis BGSC 4Y1] (ZP_04149049.1)	84 in 120 aa
205	127	-	transposase family protein	transposase family protein [Desulfosporosinus sp. OT] (ZP_08814630.1)	73 in 89 aa
206	144	-	transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 [Petrotoga mobilis SJ95] (YP_001567688.1)	45 in 141 aa
207	351	+	hypothetical protein	hypothetical protein IIE_06301 [Bacillus cereus VD045] (ZP_17566976.1)	87 in 352 aa
208	128	+	hypothetical protein	hypothetical protein IIE_06302 [Bacillus cereus VD045] (ZP_17566977.1)	54 in 127 aa
209	100	+	hypothetical protein	hypothetical protein [Bacillus cereus] (WP_016099734.1)	87 in 53 aa
210	98	+	hypothetical protein	hypothetical protein MC28_E062 [Bacillus thuringiensis MC28] (YP_006815560.1)	97 in 67 aa
211	75	+	hypothetical protein	hypothetical protein MC28_E058 [Bacillus thuringiensis MC28] (YP_006815556.1)	85 in 73 aa
212	996	-	transposase Tn3 family protein	transposase Tn3 family protein [Bacillus thuringiensis serovar finitimus YBT-020] (YP_005563651.1)	92 in 983 aa
213	280	-	integrase-recombinase protein	integrase-recombinase protein [Bacillus cereus E33L] (YP_245544.1)	90 in 280 aa

214	144	+	hypothetical protein	hypothetical protein BTG_31098 [Bacillus thuringiensis HD-771] (YP_006593824.1)	52 in 139 aa
215	54	+	hypothetical protein	hypothetical protein bthur0005_58680 [Bacillus thuringiensis serovar pakistani str. T13001] (ZP_04123905.1)	83 in 40 aa
216	70	+	hypothetical protein	hypothetical protein IEI_05877 [Bacillus cereus BAG5X2-1] (ZP_17439534.1)	52 in 63 aa
217	101	-	ArsR family transcriptional regulator	ArsR family transcriptional regulator [Bacillus thuringiensis serovar finitimus YBT-020] (YP_005569119.1)	68 in 94 aa
218	103	+	pseudogene		
219	68	+	hypothetical protein	hypothetical protein MC28_F146 [Bacillus thuringiensis MC28] (YP_006815829.1)	93 in 68 aa
220	239	-	hypothetical protein	hypothetical protein IEM_05053 [Bacillus cereus BAG6O-2] (ZP_17450491.1)	53 in 244 aa
221	366	+	response regulator aspartate phosphatase	hypothetical protein IC1_02796 [Bacillus cereus VD022] (ZP_17338319.1)	87 in 365 aa
222	80	+	hypothetical protein	hypothetical protein IEI_05823 [Bacillus cereus BAG5X2-1] (ZP_17439480.1)	78 in 74 aa
223	119	-	hypothetical protein	hypothetical protein IIE_06264 [Bacillus cereus VD045] (ZP_17566939.1)	93 in 118 aa
224	205	+	recombinase	hypothetical protein II5_06087 [Bacillus cereus MSX-A1] (ZP_17542959.1)	85 in 200 aa
225	56	+	pseudogene		
226	141	+	hypothetical protein	lipoprotein [Bacillus thuringiensis MC28] (YP_006815604.1)	96 in 141 aa
227	349	-	Fic family protein	Fic family protein [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00738430.1)	87 in 348 aa
228	326	-	hypothetical protein	hypothetical protein MC28_E157 [Bacillus thuringiensis MC28] (YP_006815655.1)	66 in 326 aa
229	404	-	lipase	lipase [Bacillus thermoamylovorans] (BAH70300.1)	55 in 387 aa
230	137	-	Cobalamin synthesis protein P47K	hypothetical protein RBTH_07780 [Bacillus thuringiensis serovar israelensis ATCC 35646] (ZP_00738433.1)	84 in 137 aa

231	103	-	Cobalamin synthesis protein P47K	hypothetical protein BTF1_32846 [Bacillus thuringiensis HD-789] (YP_006614135.1)	95 in 103 aa
232	132	-	group-specific protein	hypothetical protein pBt116 [Bacillus thuringiensis serovar israelensis] (YP_001573839.1)	93 in 132 aa
233	54	-	hypothetical protein	hypothetical protein BTF1_32336 [Bacillus thuringiensis HD-789] (YP_006614039.1)	98 in 54 aa
234	85	+	pseudogene		
235	562	-	Cry39ORF2	Cry39ORF2 protein [Bacillus thuringiensis serovar aizawai] (BAB72017.1)	83 in 558 aa
236	659	-	Cry56Ba1	Cry56Aa-like protein [Bacillus thuringiensis] (ADK38584.1)	57 in 671 aa
237	184	+	resolvase	resolvase [Bacillus cereus E33L] (YP_245576.1)	98 in 184 aa
238	467	+	Transposase for transposon Tn552	Transposase for transposon Tn552 [Bacillus cereus AH676] (ZP_04194794.1)	96 in 460 aa
239	81	+	pseudogene		
240	593	-	Fibronectin type III domain protein	hypothetical protein IIO_02806 [Bacillus cereus VD115] (ZP_17593314.1)	96 in 593 aa
241	199	-	transposon Tn552 DNA-invertase bin3	transposon Tn552 DNA-invertase bin3 [Bacillus mycoides Rock3-17] (ZP_04160431.1)	96 in 199 aa
242	982	+	Tn4652 transposase	Tn4652, transposase [Bacillus cereus Q1] (YP_002533333.1)	96 in 982 aa

LITERATURE CITED

Agaisse, H. and Lereclus, D. 1996. STAB-SD: a Shine-Dalgarno sequence in the 5' untranslated region is a determinant of mRNA stability. Mol. Microbiol. 20, 633-643.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-402.

Andrews, R.E.Jr., Bibilos, M.M. and Bulla, L.A.Jr. 1985. Protease activation of entomocidal protoxin of *Bacillus thuringiensis* subsp. *kurstaki*. Appl. Environ. Microbiol. 50, 737–732.

Andrup, L., Jensen, G.B., Wilcks, A., Smidt, L., Hoflack, L. and Mahillon, J. 2003. The patchwork nature of rolling-circle plasmids: comparison of six plasmids from two distinct *Bacillus thuringiensis* serotypes. Plasmid. 49, 205-232.

Armengol, G., Escobar, M.C., Maldonado, M.E. and Orduz, S. 2007. Diversity of Colombian strains of *Bacillus thuringiensis* with insecticidal activity against dipteran and lepidopteran insects. J. Appl. Microbiol. 102, 1, 77–88.

Aronson, A.I. 1993. The two faces of *Bacillus thuringiensis*: insecticidal proteins and post-exponential survival. Mol. Microbiol. 7, 489-496.

Aronson, A. I., Beckman, W. and Dunn. P. 1986. *Bacillus thuringiensis* and related insect pathogens. Microbiol Rev 50, 1-24.

Ash, C., Farrow, J.A., Dorsch, M., Stackebrandt, E. and Collins, M.D. 1991. Comparative analysis of *Bacillus anthracis, Bacillus cereus,* and related species on the basis of reverse transcriptase sequencing of 16S rRNA. Int. J. Syst. Bacteriol. 41, 343-346.

Barboza-Corona, J.E., Park, H.W., Bideshi, D.K. and Federici, B.A. 2012. The 60-kilodalton protein encoded by *orf2* in the *cry19A* operon of *Bacillus thuringiensis* subsp. *jegathesan* functions like a C-terminal crystallization domain. Appl. Environ. Microbiol. 78, 2005–2012.

Baum, J.A., Coyl., D.M., Gilbert., M.P., Jany., C.S. and Gawron-Burke, C. 1990. Novel cloning vectors for *Bacillus thuringiensis*. Appl. Environ. Microbiol. 56, 3420-3428.

Baum, J.A. and Gilbert., P. 1991. Characterization and comparative analysis of replication origins from three large *Bacillus thuringiensis* plasmids. J. Bacteriol. 173, 5280-5289.

Baum, J.A. and Malvar, T. 1995. Regulation of insecticidal crystal protein production in *Bacillus thuringiensis*. Mol. Microlbiol. 18, 1-12.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Wheeler, D.L. 2008. GenBank. Nucleic Acids Research. 36, 25-30.

Berry, C., Neil, S.O., Ben-Dov, E., Jones, A.F., Murphy, l., Quail, M.A., Holden, M.T.G., Harris, D., Zaritsky, A., and Parkhill, J. 2002. Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp. *israelensis*. Appl. Environ. Microbiol. 68, 5082–5095.

Boonserm, P., Davis, P., Ellar, D.J. and Li, J. 2005. Crystal structure of the mosquito -larvicidal toxin Cry4Ba and its biological implications. J. Mol. Biol. 348, 363–382.

Brantl, S., Behnke, D. and Alonso, J.C. 1990. Molecular analysis of the replication region of the conjugative *Streptococcus agalactiae* plasmid pIP501 in *Bacillus subtilis.* Comparison with plasmids pAMβ1 and pSM19035. Nucleic Acids Res. 18, 4783-4790.

Bruand, C., Lechatelier, E., Ehrlich, S.D. and Jannière, L. 1993. A fourth class of theta-replicating plasmids: The pAMβ1 family from Gram positive bacteria. Proc. Natl. Acad. Sci. USA. 90, 11668-11672.

Carroll, J., Li, J. and Ellar, D.J. 1989. Proteolytic processing of coleopteran specific delta-endotoxin produced by *Bacillus thuringiensis* var. *tenebrionis*. J. Biochem. 261, 99–105.

Carozzi, N.B., Kramer, V.C., Warren, G.W., Evola, S. and Koziel, M.G. 1991. Prediction of insecticidal activity of *Bacillus thuringiensis* strains by polymerase chain reaction product profiles. Appl. Environ. Microbiol. 57, 3057–3061.

Carlson, C.R., and Kolstø, A.B. 1993. A complete physical map of a *Bacillus thuringiensis* chromosome. J. Bacteriol. 175, 1053–1060.

Carlson, C.R., Johansen, T., Lecadet, M.M. and Kolstø, A.B. 1996. Genomic organization of the entomopathogenic bacterium *Bacillus thuringiensis* subsp. *berliner* 1715. Microbiol. 142, 1625–1634.

Carlson, C.R., Caugant, D.A. and Kolstø, A.B. 1994. Genotypic diversity among *Bacillus cereus* and *Bacillus thuringiensis* strains. Appl. Environ. Microbiol. 60, 1719–1725.

Chao, L., B. Qiyu, S. Fuping, S. Ming, H. Dafang, L. Guiming, and Y. Ziniu. 2007. Complete nucleotide sequence of pBMB67, a 67-kb plasmid from *Bacillus thuringiensis* strain YBT-1520. Plasmid. 57, 44–54.

Chattoraj, D.K. 2000. Control of plasmid DNA replication by iterons: no longer paradoxical. Mol. Microbiol. 37, 467-476.

Chen, X.J., Curtiss, A., Alcantara, E. and Dean, D.H. 1995. Mutations in domain I of *Bacillus thuringiensis* δ-endotoxin CryIAb reduced the irreversible binding of toxin to *Manduca sexta* brush border membrane vesicles. J. Biol. Chem. 270, 6412–6419.

Chen, X.J., Lee, M.K. and Dean, D.H. 1993. Site-directed mutations in a highly conserved region of *Bacillus thuringiensis* δ-endotoxin affect inhibition of

short circuit current across *Bombyx mori* midguts. Proc. Natl. Acad. Sci. USA. 90, 9041–9045.

Choi, J.Y., Kim, Y. S., Wang, Y., Tao, X.Y., Liu, Q., Roh, J.Y., Woo, S.D., Jin, B.R. and Je, Y.H. 2012. Fast and efficient generation of recombinant baculoviruses by in vitro transposition. Appl. Microbiol. Biotechnol. 96, 1353-60.

Choma, C.J. and Kaplan, H. 1991. The toxic moiety of the *Bacillus thuringiensis* protein undergoes a conformational change upon activation. Biochem. Biophys. Res. Commun. 179, 1934–1938.

Choma, C.T. and Kaplan, H. 1990. Folding and unfolding of the protoxin from *Bacillus thuringiensis*: Evidence that the toxic moiety is present in an active conformation. Biochemistry. 29, 10971–10977.

Coole, C.M. 1995. *Bacillus thuringiensis*: ecology, the significance of natural genetic modification, and regulation. World J. Microbiol. Biotechnol. 11, 471–477.

Crickmore, N., Zeigler, D.R., Schnepf, E., Van Rie, J., Lereclus, D., Baum, J., Bravo, A. and Dean, D.H. 2010. *Bacillus thuringiensis* toxin nomenclature. http://www.lifesci.sussex.ac.uk/Home/Neil_Crickmore/Bt/.

Crickmore, N., Zeigler, D.R., Schnepf, E., Van Rie, J., Lereclus, D., Baum, J., Bravo, A. and Dean, D.H. 2011. *Bacillus thuringiensis* toxin nomenclature. http://www.lifesci.sussex.ac.uk/Home/Neil_Crickmore/Bt/.

Dai, S.M. and Gill, S.S. 1993. In vitro and in vivo proteolysis of the *Bacillus thuringiensis* subsp. *israelensis* Cry IVD proteins by *Culex quinquefaciatus* larval midgut proteases. Insect Biochem. Mol. Biol. 21, 273–283.

De Barjac, H., Larget, I. and Killick-Kendrick, R. 1981. Toxicity of *Bacillus thuringiensis* var. *israelensis* serotype H14, for larvae of sandflies, the leishmaniasis vectors. Bull. Soc. Pathol. Exot. Filiales. 74, 485-489.

Dean, A.J. and Bazylinski, D.A. 1999. Genome analysis of several marine, magnetotactic bacterial strains by pulsed-field gel electrophoresis. Curr. Microbiol. 39, 219–225.

de Maagd, R.A., Bravo, A., Berry, C., Crickmore, N. and Schnepf, H.E. 2003. Structure, diversity and evolution of protein toxins from spore-forming entomopathogenic bacteria. Annu. Rev. Genet. 37, 409–433.

de Maagd, R.A., Bravo, A. and Crickmore, N. 2001. How *Bacillus thuringiensis* has evolved specific toxins to colonize the insect world. Trends Genet. 17, 193–199.

del Solar, G., Giraldo., R., Ruiz-Echevarria., M.J., Espinosa., M. and Diaz-Orejas, R. 1998. Replication and control of circular bacterial plasmids. Microbiol. Mol. Biol. Rev. 434-464.

Delcher, A. L., Harmon, D., Kasif, S., White, O. and Salzberg, S.L. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res.

27, 4636-41.

Dervyn, E., Poncet, S., Klier, A. and Rapoport, G. 1995. Transcriptional regulation of the cryIVD gene operon from *Bacillus thuringiensis* subsp. *israelensis*. J. Bacteriol. 177, 2283-91.

Ejiofor, A.O. and Johnson, T. 2002. Physiological and molecular detection of crystalliferous *Bacillus thuringiensis* strains from habitats in the South Central United States. J. Ind. Microbiol. Biotechnol. 28, 5, 284–290.

Federici, B.A. 1999. *Bacillus thuringiensis* In: Handbook of biological control. Bellows, T.S., Gordh, G., Fisher T.W. Eds. San Diego. Academic Press. 517-548.

Federici, B.A., Park, H.W. and Bideshi, D.K. 2010. Overview of the basic biology of *Bacillus thuringiensis* with emphasis on genetic engineering of bacterial larvicides for mosquito control. Open Toxinol. J. 3, 83–100.

Gamel, P.H. and Piot, J.C. 1992. Characterization and properties of a novel plasmid vector for *Bacillus thuringiensis* displaying compatibility with host plasmids. Gene. 120, 17-26.

Galitsky, N., Cody, V., Wojtczak, A., Ghosh, D., Luft, J.R., Pangborn, W. and English, L. 2001. Structure of the insecticidal bacterial delta-endotoxin Cry3Bb1 of *Bacillus thuringiensis*. Acta. Crystallogr. D. 57, 1099–1101.

Gazit, E., La Rocca, P., Sansom, M.S.P. and Shai, Y. 1998. The structure and organization within the membrane of the helices composing the pore-forming domain of *Bacillus thuringiensis* delta-endotoxin are consistent with an "umbrella-like" structure of the pore. Proc. Natl. Acad. Sci. USA. 95, 12289-12294.

Gazit, E. and Shai, Y. 1995. The Assembly and Organization of the Alpha-5 and Alpha-7 helices from the pore-forming domain of *Bacillus thuringiensis* delta-endotoxin relevance to a functional model. Journal of Biological Chemistry. 270, 2571-2578.

Gill, S.S., Cowles, E.A. and Pietrantonio, P.V. 1992. The mode of action of *Bacillus thuringiensis* endotoxins. Annu. Rev. Entomol. 37, 615–636.

González, J.M.Jr., and Carlton., B.C. 1984. A large transmissible plasmid is required for crystal toxin production in *Bacillus thuringiensis* variety *israelensis*. Plasmid. 11, 28–38.

González, J.M.Jr. and Carlton, B.C. 1980. Patterns of plasmid DNA in crystalliferous and acrystalliferous strains of *Bacillus thuringiensis*. Plasmid. 3, 92-98.

González. J.M.Jr., Dulmage, H.T. and Carlton, B.C. 1981. Correlation between specific plasmids and delta-endotoxin production in *Bacillus thuringiensis*. Plasmid. 5, 352-365.

Grochulski, P., Masson, L., Borisova, S., Pusztai-Carey, M., Schwartz, J.L.,

Brousseau, R. and Cyqler, M. 1995. *Bacillus thuringiensis* CryIA(a) insecticidal toxin: crystal structure and channel formation. J. Mol. Biol. 254, 447–464.

Guz, K., Kucinksa, J., Lonc, E. and Dorsoszkiewicz, W. 2005. Differentiated pattern of inclusions of newly isolated *Bacillus thuringiensis* strains from Silesia in Poland. Pol. J. Microbiol. 4, 263–269.

Guereca, L. and Bravo, A. 1999. The oligomeric state of *Bacillus thuringiensis* Cry toxins in solution. Biochim. Biophys. Acta. 1429, 342–350.

Heo, Y.M., Roh, J.Y., Kim, S.B., Yi, J.W., Kim, K.K., Oh, B.H. and Oh, H.T., 2012. Evaluation of the sigmoid notch involvement in the intra-articular distal radius fractures: the efficacy of computed tomography compared with plain X-ray. Clin. Orthop. Surg. 4, 83-90.

Helgason, E., Okstad, O.A., Caugant, D.A., Johansen, H.A., Fouet, A., Mock, M., Hegna, I. and Kolsto, A.B. 2000. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*-one species on the basis of genetic evidence. Appl. Environ. Microbiol. 66, 2627-2630.

Helgason, E., Caugant, D., Lecadet, M., Chen, Y., Mahillon, J., Lovgren, A., Hegna, I., Kvaløy, K. and Kolstø, A. 1998. Genetic diversity of *Bacillus cereus Bacillus thuringiensis* isolates from natural sources. Curr. Microbiol. 37, 80-87.

Hoffmaster, A.R., Ravel, J., Rasko, D.A., Chapman, G.D., Chute, M.D., Marston, C.K., De, B.K., Sacchi, C.T., Fitzgerald, C., Mayer, L.W., Maiden, M.C.J., Priest, F.G., Barker, M., Jiang, L., Cer, R.Z., Rilstone, J., Peterson, S.N., Weyant, R.S., Galloway, D.R., Read, T.D., Popovic, T. and Fraser, C.M. 2004. Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. Proc. Natl. Acad. Sci. USA. 101, 8449-8454.

Höfte, H., and Whiteley, H.R. 1989. Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiol. Rev. 53, 242-255.

Hoflack, L., Seurinck, J. and Mahillon, J. 1997. Nucleotide sequence and characterization of the cryptic *Bacillus thuringiensis* plasmid pGI3 reveal a new family of rolling circle replicons. J. Bacteriol. 179, 5000–5008.

Honee, G., Convents, D., Van Rie, J., Jansens, S., Peferoen, M. and Visser, B. 1991. The C-terminal domain of the toxic fragment of a *Bacillus thuringiensis* crystal protein determines receptor binding. Mol. Microbiol. 5, 2799-2806.

Honee, G. and Visser, B. 1993. The mode of action of *Bacillus thuringiensis* crystal proteins. Entomol. Exp. Appl. 69, 145–155.

Hu, X., Hansen, B.M., Yuan, Z., Johansen, J.E., Eilenberg, J., Hendriksen, N.B., Smidt, L. and Jensen, G.B. 2005. Transfer and expression of the mosquitocidal plasmid pBtoxis in *Bacillus cereus* group strains. FEMS.

Microbiol. Lett. 245, 239-247.

Huang, J., Guo, S., Mahillon, J., Auwera, G., Wang, Li., Han, D., Yu, Z.N. and Sun, M. 2006. Molecular characterization of a DNA fragment harboring the replicon of pBMB165 from *Bacillus thuringiensis* subsp. *tenebrionis.* BMC. Genom. 7, 260.

Hwang, S. H., Saitoh, H., Mizuki, E., Higuchi, K. and Ohba, M. 1998. A novel class of mosquitocidal delta-endotoxin, Cry19B, encoded by a *Bacillus thuringiensis* serovar *higo* gene. System. Appl. Microbiol. 21, 179-184.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W. and Hauser, L. J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. Bmc. Bioinformatics. 11.

Ibarra, J.E. and Federici, B.A. 1986. Isolation of a relatively nontoxic 65-kilodalton protein inclusion from the parasporal body of *Bacillus thuringiensis* subsp. *israelensis*. J. Bacteriol. 165, 527-533.

Ishii, T. and Ohba, M. 1993. Characterization of mosquito-specific *Bacillus thuringiensis* strains coisolated from a soil population. Syst. Appl. Microbiol. 16, 494-499.

Ito, T., Ikeya, T., Sahara, K., Bando, H. and Asano, S. 2006. Cloning and expression of two crystal protein genes, *cry30Ba1* and c*ry44Aa1*, obtained from a highly mosquitocidal strain, *Bacillus thuringiensis* subsp. *entomocidus* INA288. Appl. Environ. Microbiol. 72, 5673–5676.

Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., Bhattacharyya, A., Reznik, G., Mikhailova, N., Lapidus, A. *et al.* 2003. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. Nature, 423, 87–91.

Jensen, G.B., Wilcks, A., Petersen, S.S., Damgaard, J.J., Baum, A. and Andrup, L. 1995. The genetic basis of the aggregation system in B*acillus thuringiensis* subsp. *israelensis* is located on the large conjugative plasmid pX016. J. Bacteriol. 177, 2914–2917.

Jensen, G.B., Hansen, B.M., Eilenberg, J. and Mahillon, J. 2003. The hidden lifestyles of *Bacillus cereus* and relatives. Environm. Microbiol. 5, 631–640.

Khan, S.A. 1997. Rolling-circle replication of bacterial plasmids. Microbiol. Mol. Biol. Rev. 61, 442-455.

Koo, B.C., Kwon, M.S., Roh, J.Y., Kim, M., Kim, J.H. and Kim, T. 2012. Quantitative analysis of tetracycline-inducible expression of the green fluorescent protein gene in transgenic chickens. J. Reprod. Dev. 58, 672-677.

Kronstad, J.W., Schnepf, H.E. and Whiteley, H.R. 1983. Diversity of locations for *Bacillus thuringiensis* crystal protein genes. J. Bacteriol. 154, 419-428.

Kumar, N.S. and Venkateswerlu, G. 1997. Involvement of an endogenous metalloprotease in the activation of protoxin in *Bacillus thuringiensis* subsp.

kurstaki. Biochem. Molec. Biol. Intl. 42, 901-908.

Kumar, N.S. and Venkateswerlu, G. 1998a. Endogenous proteaseactivated 66-kDa toxin from *Bacillus thuringiensis* subsp. *kurstaki* active against *Spodoptera littoralis*. FEMS. Microbiol. Lett. 159, 113-120.

Kumar, N.S. and Venkateswerlu, G. 1998b. Intracellular proteases in sporulated *Bacillus thuringiensis* subsp. *kurstaki* and their role in protoxin activation. FEMS. Microbiol. Lett.166, 377-382.

Lecadet, M.M., Frachon, E., Dumanoir, V.C., Ripouteau, H., Hamon, S., Laurent, P. and Thiery, I. 1999. Updating the H-antigen classification of *Bacillus thuringiensis*. J. Appl. Microbiol. 86, 660-672.

Lee, H.K. and Gill, S.S. 1997. Molecular cloning and characterization of a novel mosquitocidal protein gene from *Bacillus thuringiensis* subsp. *fukuokaensis*. Appl. Environ. Microbiol. 63, 4664-4670.

Lee, I.H., Je, Y.H., Chang, J.H., Roh, J.Y., Oh, H.W., Lee, S.G., Shin, S.C. and Boo, K.S. 2001. Isolation and characterization of a *Bacillus thuringiensis* ssp. *kurstaki* strain toxic to *Spodoptera exigua* and *Culex pipiens*. Current. Microbiol. 43, 284-287.

Lenane, I.J., Bagnall, N.H., Josh, P.F., Pearson, R.D., Akhurst, R.J. and Kotze, A.C. 2008. A pair of adjacent genes, *cry5Ad* and *orf2-5Ad*, encode the typical N- and C-terminal regions of a Cry5Aδ-endotoxin as two separate proteins in *Bacillus thuringiensis* strain L366. FEMS. Microbiol. Lett. 278, 115–120.

Lereclus, D., Arantes, Q., Chaufaux, J., and Lecadet, M. 1989. Transformation and expression a cloned delta-endotoxin gene in *Bacillus thuringiensis*. FEMS. Microbiol. Lett. 51, 211–217.

Li, M.S., Je, Y.H., Lee, I.H., Chang, J.H., Roh, J.Y., Kim, H.S., Oh, H.W. and Boo, K.S. 2002. Isolation and characterization of a strain of *Bacillus thuringiensis* ssp. *kurstaki* containing a new delta-endotoxin gene. Curr. Microbiol. 45, 299-302.

Li, Z., Hiasa, H. and DiGate, R. 2005. *Bacillus cereus* DNA topoisomerase I and III alpha: purification, characterization and complementation of *Escherichia coli* TopoIII activity. Nucleic Acids Res. 33, 5415-5425.

Lukashin, A.V. and Borodovsky, M. 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26, 1107-1115.

Mahillon, J., Rezsohazy, R., Hallet, B. and Delcour, J. 1994. IS231 and other *Bacillus thuringiensis* transposable elements: a review. Genetica. 93,13-26.

Manasherob, R., Zaritsky, A., Ben-Dov, E., Saxena, D., Barak, Z. and Einav, M. 2001. Effect of accessory proteins P19 and P20 on cytolytic activity of Cyt1Aa from *Bacillus thuringiensis* subsp. *israelensis* in *Escherichia coli*. Curr. Microbiol. 43, 355-364.

Martin, P.A. and Travers, R.S. 1989. Worldwide abundance and distribution of

Bacillus thuringiensis isolates. Appl. Environ. Microbiol. 55, 2437–2442.

Nickerson, K.W. and Bulla, L.A.Jr. 1974. Physiology of sporeforming bacteria associated with insects: minimal nutritional requirements for growth, sporulation, and parasporal crystal formation of *Bacillus thuringiensis*. Appl. Microbiol. 28, 124-128.

Nishimoto, T., Yoshisue, H., Ihara, K., Sakai, H. and Komano, T. 1994. Functional analysis of block 5, one of the highly conserved amino acid sequences in the 130-kDa CryIVA protein produced by *Bacillus thuringiensis* subsp. *israelensis*. FEBS. Lett. 348, 249-254.

Noguera, P.A. and Ibarra, J.E. 2010. Detection of new *cry* genes of *Bacillus thuringiensis* by use of a novel PCR primer system. Appl. Environ. Microbiol. 76, 6150–6155.

Ohba, M. and Aizawa, K. 1978. Serological identification of *Bacillus thuringiensis* and related bacteria isolated in Japan. J. Invertebr. Pathol. 32, 303-309.

Ohba, M. and Aizawa, K. 1989. New flagellar (H) antigenic subfactors in *Bacillus thuringiensis* H serotype 3 with description of two new subspecies, *Bacillus thuringiensis* subsp. *sumiyoshiensis* (H serotype 3a:3d) and *Bacillus thuringiensis* subsp. *fukuokaensis* (H serotype 3a:3d:3e). J. Invertebr. Pathol. 54, 208-212.

Ohba, M., Mizuki, E. and Uemori, A. 2009. Parasporin, a new anticancer protein group from Bacillus thuringiensis. Anticancer Res., 29, 427–433.

Ohgushi, A., Saitoh, H., Wasano, N., Uemori, A. and Ohba, M. 2005. Cloning and characterization of two novel genes, *cry24B* and *s1orf2*, from a mosquitocidal strain of *Bacillus thuringiensis* serovar *sotto*. Curr. Microbiol. 51, 131–136.

Park, H.W., Ge, B., Bauer, L.S. and Federici, B.A. 1998. Optimization of Cry3A yields in *Bacillus thuringiensis* by use of sporulation-dependent promoter in combination with the STAB-SD mRNA sequence. Appl. Environ. Microbiol. 64, 3932–3938.

Porcar, M. and Juárez-Pérez, V.M. 2003. PCR-based identification of *Bacillus thuringiensis* pesticidal crystal genes. FEMS. Microbiol. Rev. 26, 5, 419–432.

Rasko, D. A., Altherr, M. R., Han, C. S. and Ravel, J. 2005. Genomics of the *Bacillus cereus* group of organisms. FEMS. Microbiol. Rev. 29, 303-329.

Read, T.D., Peterson, S.N., Tourasse, N., Baillie, L.W., Paulsen,I .T., Nelson ,K.E., Tettelin, H., Fouts, D.E., Eisen, J.A., Gill, S.R. *et al.* 2003. The genome sequence of *Bacillus anthracis Ames* and comparison to closely related bacteria. Nature, 423, 81–86.

Reyes-Ramirez, A. and Ibarra, J.E. 2008. Plasmid patterns of *Bacillus thuringiensis* type strains. Appl. Environ. Microbiol. 74, 125-129.

Roh, J.Y., Choi, J.Y., Jin, B.R. and Je, Y.H. 2007. *Bacillus thuringiensis* as a specific, safe, and effective tool for insect pest control. J. Microbiol.

Biotechnol. 17, 547–559.

Roh, J.Y., Liu, Q., Lee, D. W., Tao, X., Wang, Y., Shim, H.J., Choi, J.Y., Seo, J.B., Ohba, M., Mizuki, E. and Je, Y.H. 2009. *Bacillus thuringiensis* serovar *mogi* (flagellar serotype 3a3b3d), a novel serogroup with a mosquitocidal activity. J. Invertebr. Pathol. 102, 266-268.

Roh, J.Y., Park, H.W., Jin, B.R., Kim, H.S., Yu, Y.M. and Kang, S.K. 1996. Characterization of novel non-toxic *Bacillus thuringiensis* isolates from Korea. Lett. In. Appl. Microbiol. 23, 249-252.

Rosso, M. and Delécuse, A. 1997. Contribution of the 65-kilodalton protein encoded by the cloned gene *cry19A* to the mosquitocidal activity of *Bacillus thuringiensis* subsp. *jegathesan*. Appl. Environ. Microbiol. 63, 4449–4455.

Russell, R.M., Robertson, J.L. and Sauvin, S.E. 1977. POLO: A new computer program for probit analysis. Bull. Entomol. Soc. Am. 23, 209-213.

Rukmini, V., Reddy, C.Y. and Venkateswerlu, G. 2000. *Bacillus thuringiensis* crystal δ-endotoxin: role of proteases in the conversion of protoxin to toxin. Biochim. 2, 109–116.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A. and Barrell, B. 2000. Artemis: sequence visualization and annotation. Bioinfor. 16, 944-945.

Saitoh, H., Hwang, S.H., Park, Y.S., Higuchi, K., Mizuki, E. and Ohba, M. 2000. Cloning and characterization of a *Bacillus thuringiensis* serovar *higo* gene encoding a novel class of the delta-endotoxin protein, Cry27Aa, specifically active on the *Anopheles* mosquito. System. Appl. Microbiol. 23, 25-30.

Salehi, G.R., Seifinejad, A., Saeedizadeh, A., Nazarian, A., Yousefloo, M., Soheilivand, S., Mousivand, M., Jahangiri, R., Yazdani, M., Maali, A.R. and Akbari, S. 2008. Molecular detection of nematicidal crystalliferous *Bacillus thuringiensis* strains of Iran and evaluation of their toxicity on free living and plant parasitic nematodes. Can. J Microbiol. 10, 812–822.

Schneph, H.E., Whiteley, H.R. 1981. Cloning and expression of the *Bacillus thuringiensis* crystal protein gene in *Escherichia col*i. Proc. Natl. Acad. Sci. USA 78, 2893-2897.

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D.R. and Dean, D.H. 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol Mol Biol Rev. 62, 775-806.

Sezen, K., Kati, H., Muratoglu, H. and Demirbag, Z. 2010. Characterization and toxicity of *Bacillus thuringiensis* strains from hazelnut pests and fields. Pest Manag. Sci. 66, 543–548.

Sutanto, Y., Shoemaker, N.B., Gardner, J.F. and Salyers, A.A. 2002. Characterization of Exc, a novel protein required for the excision of Bacteroides conjugative transposon. Mol. Microbiol., 46, 1239–1246.

Swinfield, T., Oultram, J.D., Thompson, D.E., Brehm, J.K. and Minton, N.P. 1990. Physical characterization of the replication region of the *Streptococcus faecalis* plasmid pAMβ1. Gene. 87, 79-90.

Tang, M., Bieshi, D.K., Park, H.W. and Federici, B.A. 2006. Minireplicon from pBtoxis of *Bacillus thuringiensis* subsp. *isrelensis*. Appl. Environ. Microbiol. 72, 6984-6954.

Thomas, W.E. and Ellar, D.J. 1983. *Bacillus thuringiensis* var *israelensis* crystal delta-endotoxin: effects on insect and mammalian cells in vitro and in vivo. J. Cell Sci. 60, 181-197.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Thorne, L., Garduno, F., Thompson, T., Decker, D., Zounes, M.A., Wild, M., Walfield, A.M. and Pollock, T.J. 1986. Structural similarity between the lepidoptera- and diptera-specific insecticidal endotoxin genes of *Bacillus thuringiensis* subsp. "*kurstaki*" and "*israelensis*". J. Bacteriol. 166, 801–811.

Tinsley, E., Naqvi, A., Bourgogne, A., Koehler, T.M. and Khan, S.A. 2004. Isolation of a minireplicon of the virulence plasmid pXO2 of *Bacillus anthracis* and characterization of the plasmid-encoded RepS replication protein. J. Bacteriol. 186, 2717-2723.

Tinsley, E. and Khan, S.A. 2006. A novel FtsZ-like protein is involved in replication of the anthrax toxin-encoding pXO1 plasmid in *Bacillus anthracis*. J. Bacteriol. 188, 2829-2835.

Vaughan, S., Wickstead, B., Gull, K. and Addinall, S.G. 2004. Molecular evolution of FtsZ protein sequences encoded within the genomes of Archaea, Bacteria, and Eukaryota. J. Mol. Evol. 58, 19-39.

Vontersch, M.A., Slatin, S.L., Kulesza, C.A. and Eglish, G.H. 1994. Membrane permeabilising activities of *Bacillus thuringiensis* Cry IIIB2 and Cry IIIB2 domain I peptide. Appl Environ Microbiol. 60, 3711–3717.

Wilcks, A., Smidt., L., Okstad., O. A., Kolsto., A. B., Mahillon., J. and Andrup, L. 1999. Replication mechanism and sequence analysis of the replicon of pAW63, a conjugative plasmid from *Bacillus thuringiensis*. J. Bacteriol. 181, 3193 -3200.

Wu, D. and Federici, B.A., 1993. A 20-kilodalton protein preserves cell viability and promotes cyta crystal-formation during sporulation in *Bacillus thuringiensis*. J. Bacteriol. 175, 5276-5280.

Xu, Y.N., Uhm, S.J., Koo, B.C., Kwon, M.S., Roh, J.Y., Yang, J.S., Choi, H.Y., Heo, Y.T., Cui, X.S., Yoon, J.H., Ko, D.H., Kim, T. and Kim, N.H. 2013. Production of transgenic Korean native cattle expressing enhanced green

fluorescent protein using a FIV-based lentiviral vector injected into MII oocytes. J. Genet. Genomics. 40, 37-43.

Yamagiwa, M., Esaki, M., Otake, K., Inagaki, M., Komano, T., Amachi, T. and Sakai, H. 1999. Activation process of dipteran-specific insecticidal protein produced by *Bacillus thuringiensis* subsp. *israelensis*. Appl. Environ. Microbiol. 65, 3464-3469.

Yamamoto, T. and Powell, G.K. 1993. *Bacillus thuringiensis* crystal proteins: recent advances in understanding its insecticidal activity. In: Kim, L. (Ed.), Advanced Engineered Pesticides. Marcel Dekker Inc., New York, pp. 3–42.

Zhao, X., Zhang, Z., Yan, J. and Yu, J. 2007. GC content variability of eubacteria is governed by the pol III alpha subunit. Biochem. Biophys. Res. Commun. 356, 20-25.

Zhao, X.Q., Hu, J.F. and Yu, J. 2006. Comparative analysis of eubacterial DNA polymerase III alpha subunits. Genomics Proteomics Bioinfomatics. 4, 203-211.

모기 살충성 strain, *Bacillus thuringiensis* subspecies *mogi* 의 분자 생물학적 특성 연구

서울대학교

농생명공학부 곤충학전공

유금

초 록

국내 문경지역에서 수집한 낙엽으로부터 분리한 *Bacillus thuringiensis* subsp. *mogi* 균주는 plasmid 상에 곤충병원성과 관련된 다양한 유전자를 가지고 있다. 따라서 본 연구에서는 *B. thuringiensis* subsp. *mogi* 균주의 생물학적 특성을 구명하고, 전체 genome 염기서열 및 유전자 구조를 분석하며, plasmid 상에 위치한 새로운 cry 유전자의 발현에 대한 분자생물학적 특성을 조사하고자 하였다.

55 개 type-strain 의 편모 항혈청을 이용하여 편모항원성을 검정한 결과 *B. thuringiensis* subsp. *mogi* 균주는 3a3b3d의 새로운 serogroup인 것으로 나타났으며, 이 결과를 바탕으로 serovar *mogi* 로 명명하였다. 이러한 *B. thuringiensis* subsp. *mogi* 균주는 *Culex pipiens molestus* 와 *Culex pipiens pallens* 등 파리목 유충에 대해서는 살충활성을 보였지만 나비목 유충에 대해서는 살충활성을 보이지 않았다. 또한, *B. thuringiensis* subsp. *mogi* 균주는 3 개의 난형의 parasporal crystal 이 하나의 envelope 에 둘러싸인 형태의 inclusion body 를 생성하였으며, SDS-PAGE 를 수행한 결과 이들 parasporal crystal 은 30~75 kDa 정도의 분자량을 가진 여러개의 단백질로 이루어져 있는 것으로

나타났다. 이들 단백질 band 에 대하여 nano-LC-ESI-IT MS 분석을 수행한 결과, Cry27Aa, Cry39ORF2 및 Cry20-like 의 putative peptides 인 것으로 확인되었다. 한편, H3 serotype 에 속하는 기존의 *B. thuringiensis* 균주들은 복잡한 plasmid profile 을 보이는데 비해 *B. thuringiensis* subsp. *mogi* 균주는 30 MDa 이상의 megaplasmid 만 보유하고 있는 것으로 나타났다.

B. thuringiensis subsp. *mogi* 균주의 전체 genome 은 약 6.0 Mb 였으며, 5,652 개의 ORF 를 coding 하고 있는 circular chromosome (약 5.4 Mb)과 두개의 megaplasmid, pMOGI364 (364,564 bp) 및 pMOGI222 (222,348 bp) 등 총 세 개의 replicon 으로 이루어져 있었고, 이들 replicon 의 G+C contents 는 31.3~35.2%였다. 두 개의 megaplasmid 상에는 모두 17 개의 병원성 관련 cry 유전자가 존재하는 것으로 분석되었으며, 이중 6 개의 유전자 (*cry19Bb1*, *cry73Aa* 과 *cry4orf2* operon, *cry20Bb1, cry27Ab1, cry4Aa, cry56Ba1* 과 *cry39orf2* operon)가 기존의 *cry* 유전자에서 보고된 세 개의 domain 을 모두 가지는 것으로 보아 실질적인 살충활성을 가질 것으로 예상되었다.

앞에서 살충활성을 가질 것으로 예상된 6 개의 새로운 *cry* 유전자에 대하여 정량 PCR (qPCR) 을 수행한 결과, 이들 유전자 모두가 *B. thuringiensis* subsp. *mogi* 균주 내에서 정상적으로 transcription 이 되는 것을 확인할 수 있었다. 이들 *cry* 유전자의 발현 특성을 알아보기 위하여 자신의 promoter 의 조절 하에서 *Escherichia coli-B. thuringiensis* shuttle vector 인 pHT1K 에 cloning 하고 acrystalliferous *B. thuringiensis* Cry-B 균주에 도입한 결과, *cry20Bb1* 과 *cry56Ba1* operon 이 형질전환된 균주에서만 wild-type *B.*

thuringiensis subsp. *mogi* 균주에서보다 그 크기는 작지만 crystal 을 형성하였으며, crystal 을 형성한 경우에만 모기 유충에 살충활성을 보였다. 한편, *cry56Ba1* operon 에서 *cry39orf2* 의 역할을 알아보기 위하여 이들 유전자를 STAB-SD sequence 와 강력한 chimeric *cyt1Aa* promoter 를 가진 over-expression vector 인 p1KSD 에 cloning 하고 acrystalliferous *B. thuringiensis* Cry-B 균주에 도입하였다. 그 결과, intact 한 operon 구조뿐만 아니라 *cry39orf2* 만 발현시켜도 crystal 을 형성하는 것을 확인할 수 있었다. 이러한 결과는 Cry39ORF2 가 기존의 Cry 단백질에서 structural region 으로 보고된 C-말단의 역할을 하여 Cry59Ba1 의 crystallization 에 관여한다는 것을 암시하였다.

검 색 어 : *B.thuringiensis,* ovoidal-shaped crystals, mosquitocidal, full genome sequence, three-domain *cry* gene, over-expression

학번: 2007-23602

감사의 글

2013년의 겨울은 기다림 속에서 맞이하였습니다. 서울대학교에서 보낸 6년반의 세월동안 좋으신 분들을 많이 만날수 있었던 것은 저에게는 큰 행운이었습니다. 박사학위 논문이 끝나는 이 시점에서, 저는 저에게 도움을 주신 많은 고마운 분들께 깊은 감사의 인사를 올리고 싶습니다.

먼저, 저의 지도교수님이신 제연호 교수님께 가장 진심 어린 경의와 고마움을 전하고 싶습니다. 저는 교수님의 학생이 된 것을 매우 영광으로 생각합니다. 교수님은 일상생활과 학업에서 모두 주도면밀한 관심을 주셨는바, 교수님의 은혜는 평생 가슴에 새기도록 하겠습니다. 학문을 대하는 교수님의 태도, 넓은 흉금, 정직한 인격, 긍정적인 에너지는 제가 평생 따라 배워야 하는 본보기입니다. 언제 어디서나 교수님 생각만 하면 저의 마음속에는 더없는 자부감이 생깁니다. 저도 교수님과 같은 사람으로 거듭나도록 노력할 것입니다. 재학 동안, 제가 교수님께 끼친 폐에 대하여 매우 미안한 마음이 있으며 동시에 교수님께서 너그럽게 헤아려 주시고 이해해 주신데 대해 고마울 따름입니다.

또한 저의 눈문심사위원이신 안용준 교수님, 이승환 교수님, 국립보건연구원 노종열 박사님, 전북대학교 김재수 박사님께 감사드립니다. 교수님들은 바쁘신 와중에도 보귀한 시간을 할애하여 저의 논문을 읽어주시고 수정해 주셨으며 많은 소중한 견해를 제기하여 주셨습니다.

그리고 학과의 이준호 교수님, 이시혁 교수님, 이광범 교수님한테 고마움을 전하고 싶습니다. 여러해 동안 교수님들의 생동감있고 풍부한 강의를

감명깊게 들었으며 교수님들로 하여 서울대학교 곤충전공은 현재와 같이 나날이 발전할 수 있을 것이라 생각됩니다.

심혈을 기울여 저의 실험을 첫번째로 도와주시고 *Bacillus thuringiensis* 의 세계로 이끌어주신 노종열 박사님한테 감사의 인사를 올립니다. 박사님으로 부터 저는 하나의 완전한 실험방안의 건립방법을 배웠으며 과학연구 과정에서 즐거움을 찾는 방법을 배웠습니다.

그리고 실험실의 학우들-최재영 박사님, 김우진 박사님, 후배 박종빈, 이주현, 김송은, 이석희, 안샛별, 방영, 김종훈에게 고마움을 전하고 싶습니다. 이들은 저에게 여러 면에서 도움을 주었고 실험실에서 함께 제가 발전하는 모습을 지켜봐 주었습니다. 우리의 "행복한 대가정" 성원들이 영원히 행복하길 바랍니다.

또한 임재윤 박사님의 아낌없는 도움에 감사를 드립니다.

동시에 한국땅에 밟은 후 첫번째로 친절하게 다가왔고, 항상 오빠처럼 저를 보살펴 준 Wang Yong형에게 고마우을 전하고, Tao Xueying과는 함께 5년이라는 시간을 보냈는데 우리의 방황, 고통, 기쁨은 서로에게 제일 좋은 격려가 되었습니다. Tao Xueying와 Tao Xueying의 아이가 늘 건강하고 행복하길 바랍니다.

그리고 잊지못할 나날들을 함께 보낸 Lin Yang, Cheng Jing, Liu Yang, Gao Yaqian, Jin Yinhua , Gao Zhe, Fu zhenli, Wang zhangqian, Wang xue, Zhang Xiangmei, He Hai, Cheng Yanpin, Liu Xiaoshan, Wu Lili, Song Xiangqing , Shi Siyu , Jing Liang, 등 나의 친구들에게 감사드립니다. 친구들

은 저한테 많은 관심과 격려를 주었고 그들로인해 나의 한국 생할은 더욱 즐겁고 따뜻했습니다. 친구들에게 많은 것을 배웠으며 함께 있었던 나날들은 참으로 즐거운 시간들이었습니다.

마지막으로 저의 아버지, 어머니, 여동생, 남동생한테 감사의 인사를 올립니다. 그들은 저에게 아낌없는 사랑을 주었고 저의 학업을 위해 거대한 희생과 노력을 대가로 하였다는 것을 잘 알고 있지만 저는 아직 그 은혜를 갚지 못했습니다.

저의 가족들께, 그리고 저를 관심하여 주신 모든 여러분께 다시한번 무한한 축복을 드립니다. 여러분의 행복과 건강을 기원합니다.